

## 特許協力条約

PCT

## 国際予備審査報告

REC'D 21 MAY 1999

WIPO PCT

(法第12条、法施行規則第56条)  
[PCT36条及びPCT規則70]

出願人又は代理人 の書類記号 98215M	今後の手続きについては、国際予備審査報告の送付通知（様式PCT/IPEA/416）を参照すること。	
国際出願番号 PCT/JP98/04457	国際出願日 (日.月.年) 02.10.98	優先日 (日.月.年) 02.10.97
国際特許分類 (IPC) Int. Cl. G01N33/68, G06F17/00, G06F15/18, C07K14/00		
出願人（氏名又は名称） 板井昭子		

1. 国際予備審査機関が作成したこの国際予備審査報告を法施行規則第57条（PCT36条）の規定に従い送付する。

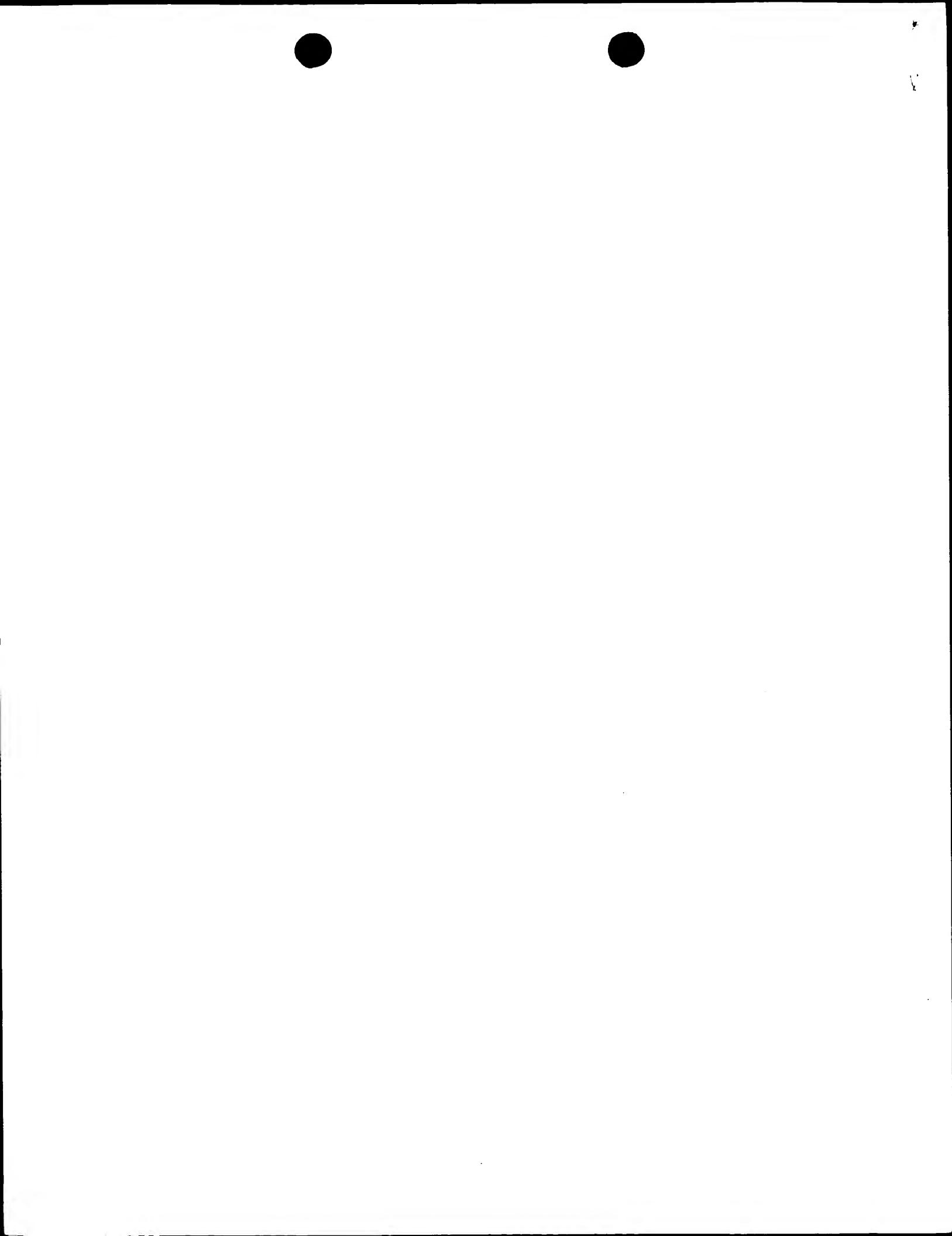
2. この国際予備審査報告は、この表紙を含めて全部で 4 ページからなる。

この国際予備審査報告には、附属書類、つまり補正されて、この報告の基礎とされた及び／又はこの国際予備審査機関に対して訂正を含む明細書、請求の範囲及び／又は図面も添付されている。  
(PCT規則70.16及びPCT実施細則第607号参照)  
この附属書類は、全部で        ページである。

3. この国際予備審査報告は、次の内容を含む。

I  国際予備審査報告の基礎  
II  優先権  
III  新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成  
IV  発明の單一性の欠如  
V  PCT35条(2)に規定する新規性、進歩性又は産業上の利用可能性についての見解、それを裏付けるための文献及び説明  
VI  ある種の引用文献  
VII  国際出願の不備  
VIII  国際出願に対する意見

国際予備審査の請求書を受理した日 02.10.98	国際予備審査報告を作成した日 30.04.99
名称及びあて先 日本国特許庁 (IPEA/JP) 郵便番号 100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員） 2J 9015 龜田宏之 電話番号 03-3581-1101 内線 3252



## I. 国際予備審査報告の基礎

1. この国際予備審査報告は下記の出願書類に基づいて作成された。(法第6条(PCT14条)の規定に基づく命令に応答するために提出された差し替え用紙は、この報告書において「出願時」とし、本報告書には添付しない。PCT規則70.16, 70.17)

 出願時の国際出願書類

<input type="checkbox"/> 明細書 第 _____ ページ	出願時に提出されたもの
明細書 第 _____ ページ	国際予備審査の請求書と共に提出されたもの
明細書 第 _____ ページ	付の書簡と共に提出されたもの
<input type="checkbox"/> 請求の範囲 第 _____ 項	出願時に提出されたもの
請求の範囲 第 _____ 項	PCT19条の規定に基づき補正されたもの
請求の範囲 第 _____ 項	国際予備審査の請求書と共に提出されたもの
請求の範囲 第 _____ 項	付の書簡と共に提出されたもの
<input type="checkbox"/> 図面 第 _____ ページ/図	出願時に提出されたもの
図面 第 _____ ページ/図	国際予備審査の請求書と共に提出されたもの
図面 第 _____ ページ/図	付の書簡と共に提出されたもの
<input type="checkbox"/> 明細書の配列表の部分 第 _____ ページ	出願時に提出されたもの
明細書の配列表の部分 第 _____ ページ	国際予備審査の請求書と共に提出されたもの
明細書の配列表の部分 第 _____ ページ	付の書簡と共に提出されたもの

2. 上記の出願書類の言語は、下記に示す場合を除くほか、この国際出願の言語である。

上記の書類は、下記の言語である \_\_\_\_\_ 語である。

- 国際調査のために提出されたPCT規則23.1(b)にいう翻訳文の言語
- PCT規則48.3(b)にいう国際公開の言語
- 国際予備審査のために提出されたPCT規則55.2または55.3にいう翻訳文の言語

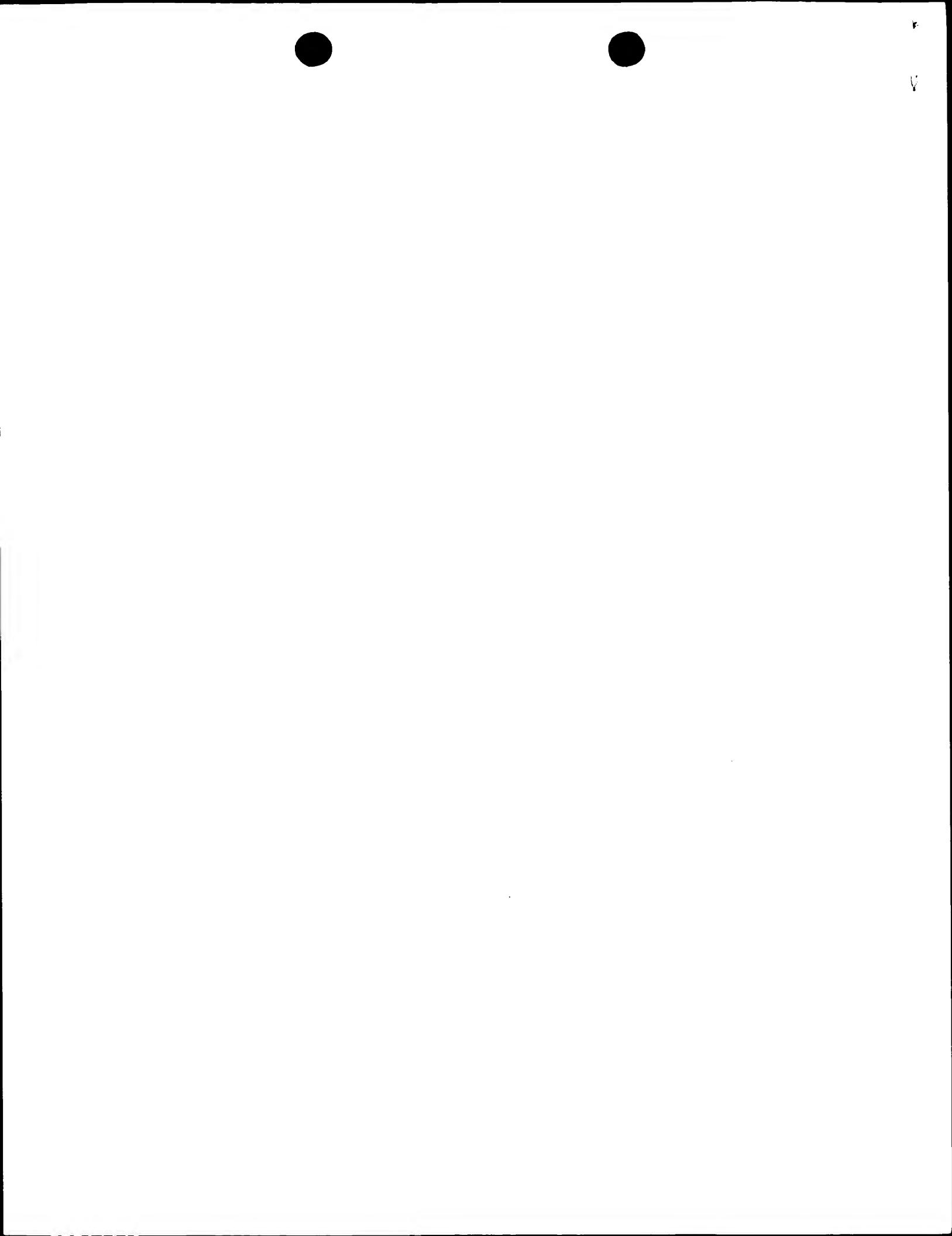
3. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき国際予備審査報告を行った。

- この国際出願に含まれる書面による配列表
- この国際出願と共に提出されたフレキシブルディスクによる配列表
- 出願後に、この国際予備審査(または調査)機関に提出された書面による配列表
- 出願後に、この国際予備審査(または調査)機関に提出されたフレキシブルディスクによる配列表
- 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった
- 書面による配列表に記載した配列とフレキシブルディスクによる配列表に記録した配列が同一である旨の陳述書の提出があった。

4. 挽正により、下記の書類が削除された。

- 明細書 第 \_\_\_\_\_ ページ
- 請求の範囲 第 \_\_\_\_\_ 項
- 図面 図面の第 \_\_\_\_\_ ページ/図

5.  この国際予備審査報告は、補充欄に示したように、挽正が出願時における開示の範囲を越えてされたものと認められるので、その挽正がされなかったものとして作成した。(PCT規則70.2(c) この挽正を含む差し替え用紙は上記1.における判断の際に考慮しなければならず、本報告に添付する。)



## III. 新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成

1. 次に関して、当該請求の範囲に記載されている発明の新規性、進歩性又は産業上の利用可能性につき、次の理由により審査しない。

国際出願全体

請求の範囲 11

理由：

この国際出願又は請求の範囲 11 は、国際予備審査をすることを要しない次の事項を内容としている（具体的に記載すること）。

データベースは情報の単なる提示に相当する。

明細書、請求の範囲若しくは図面（次に示す部分）又は請求の範囲 の記載が、不明確であるため、見解を示すことができない（具体的に記載すること）。

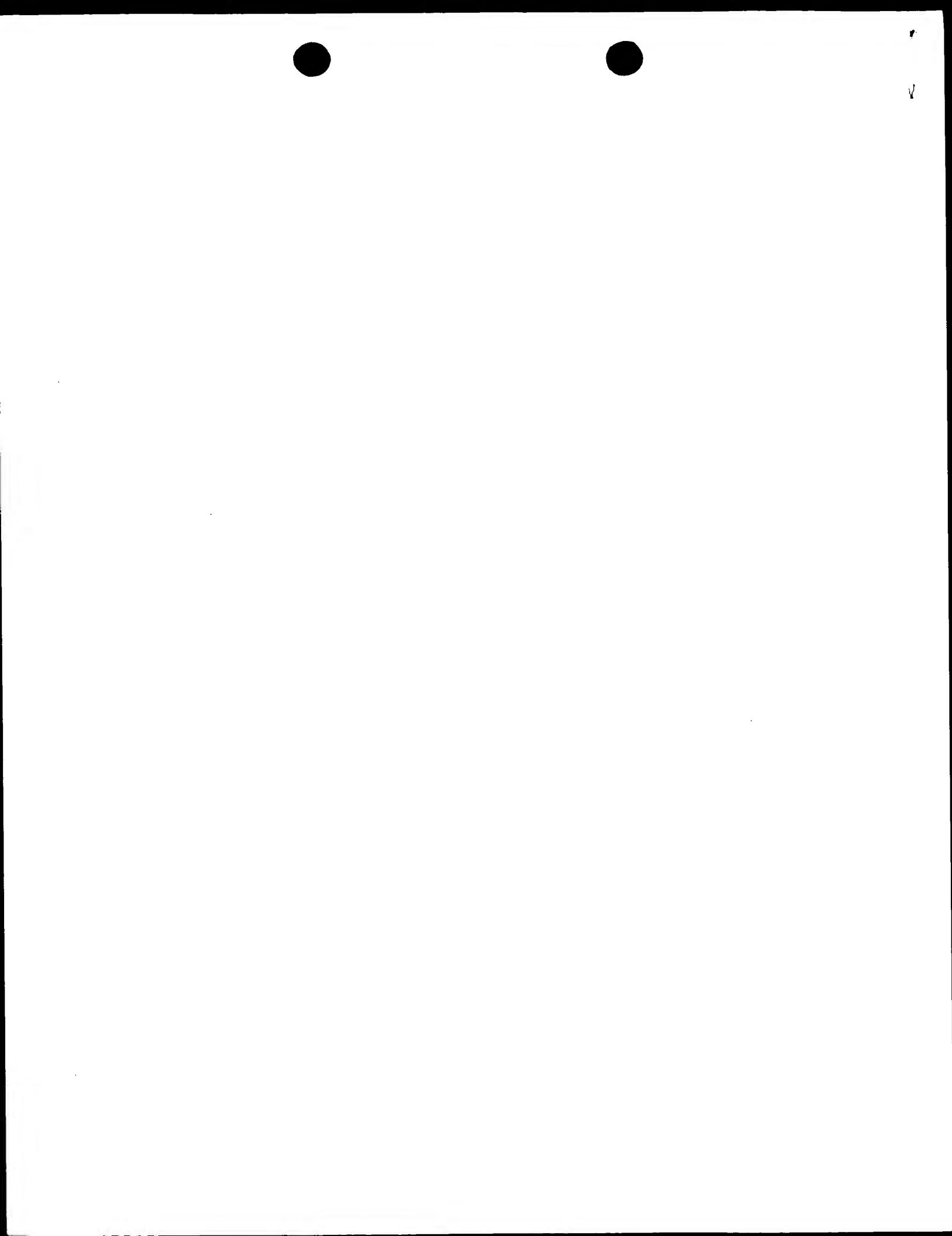
全部の請求の範囲又は請求の範囲 が、明細書による十分な裏付けを欠くため、見解を示すことができない。

請求の範囲 について、国際調査報告が作成されていない。

2. ヌクレオチド又はアミノ酸の配列表が実施細則の附属書C（塩基配列又はアミノ酸配列を含む明細書等の作成のためのガイドライン）に定める基準を満たしていないので、有効な国際予備審査をすることができない。

書面による配列表が提出されていない又は所定の基準を満たしていない。

フレキシブルディスクによる配列表が提出されていない又は所定の基準を満たしていない。



## V. 新規性、進歩性又は産業上の利用可能性についての法第12条 (PCT35条(2)) に定める見解、それを裏付ける文献及び説明

## 1. 見解

新規性 (N)

請求の範囲 1 ~ 10 有  
請求の範囲 \_\_\_\_\_ 無

進歩性 (I S)

請求の範囲 \_\_\_\_\_ 有  
請求の範囲 1 ~ 10 無

産業上の利用可能性 (I A)

請求の範囲 1 ~ 10 有  
請求の範囲 \_\_\_\_\_ 無

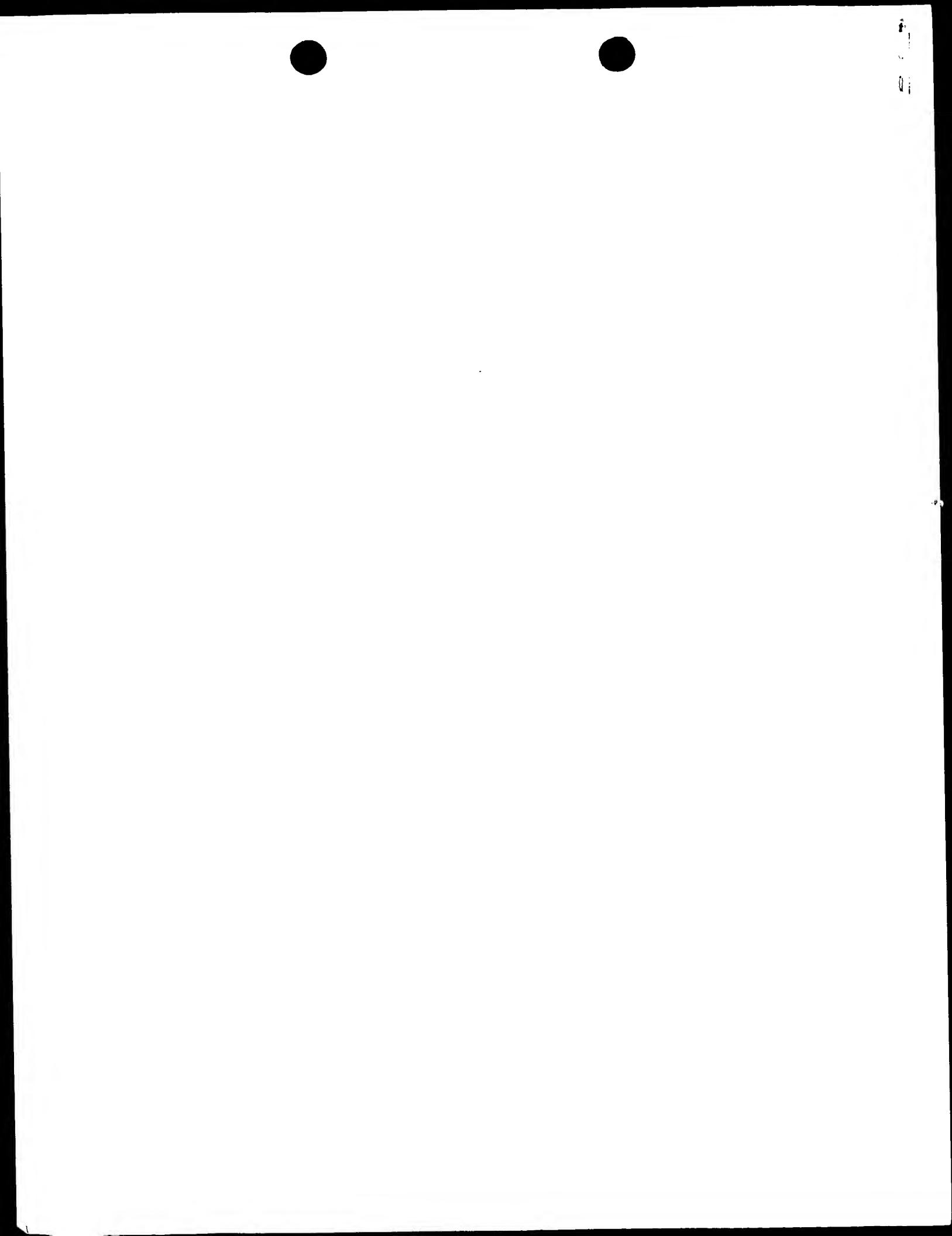
## 2. 文献及び説明 (PCT規則70.7)

請求の範囲 1 ~ 10

文献1: J P, 7-206894, A (アドバンスドテクノロジーインスティチュート株式会社) 8. 8月. 1995 (08. 08. 95)

文献1には、蛋白質の立体構造の予測に際し、立体構造既知の蛋白質で相同性が高いものを選択して検討することが記載されている。その場合に、極性度やpK値等の物理化学的パラメータについて検討することも記載されている。

パラメータとしてアミノ酸残基の側鎖の疎水性や親水性を用いることは当業者にとって自明である。



## PATENT COOPERATION TREATY

PCT

NOTIFICATION OF RECEIPT OF  
RECORD COPY

(PCT Rule 24.2(a))

From the INTERNATIONAL BUREAU

To:

IMAMURA, Masazumi  
KRF Building  
5th floor  
5-5, Kyobashi 1-chome 1998.11.17  
Chuo-ku  
Tokyo 104-0031

**RECEIVED**  
**SI&K**

SHIOZAWA, IMAMURA  
& KAMATA

Date of mailing (day/month/year) 16 October 1998 (16.10.98)	IMPORTANT NOTIFICATION
Applicant's or agent's file reference 98215M	International application No. PCT/JP98/04457

The applicant is hereby notified that the International Bureau has received the record copy of the international application as detailed below.

Name(s) of the applicant(s) and State(s) for which they are applicants:

ITAI, Akiko (all designated States)

International filing date : 02 October 1998 (02.10.98)  
 Priority date(s) claimed : 02 October 1997 (02.10.97)  
 Date of receipt of the record copy by the International Bureau : 16 October 1998 (16.10.98)

List of designated Offices :

EP :AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE  
 National :JP,US

## ATTENTION

The applicant should carefully check the data appearing in this Notification. In case of any discrepancy between these data and the indications in the international application, the applicant should immediately inform the International Bureau.

In addition, the applicant's attention is drawn to the information contained in the Annex, relating to:

- time limits for entry into the national phase
- confirmation of precautionary designations
- requirements regarding priority documents

A copy of this Notification is being sent to the receiving Office and to the International Searching Authority.

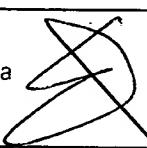
The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

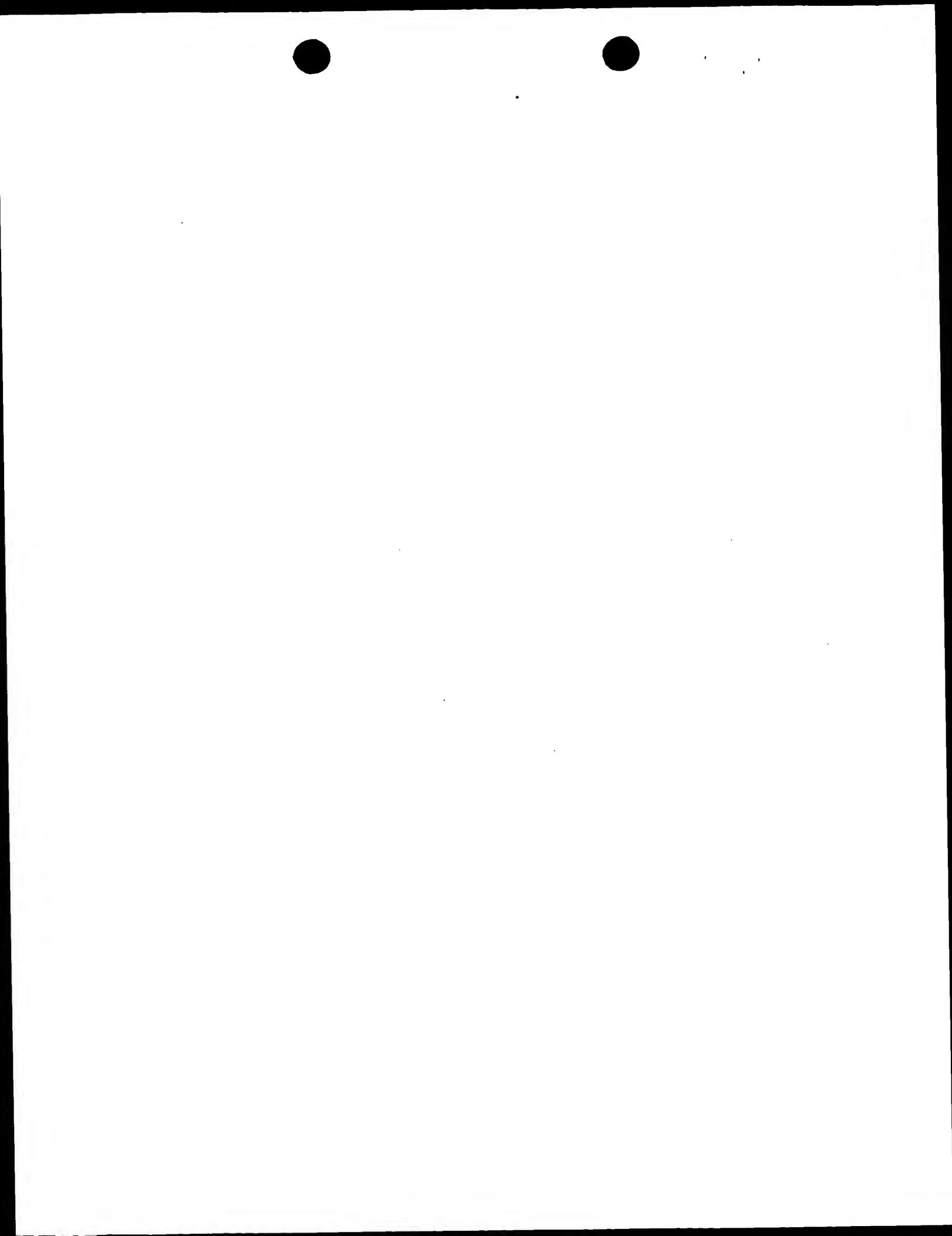
Facsimile No. (41-22) 740.14.35

Authorized officer:

K. Takeda

Telephone No. (41-22) 338.83.38





## PATENT COOPERATION TREATY

PCT

NOTIFICATION CONCERNING  
SUBMISSION OR TRANSMITTAL  
OF PRIORITY DOCUMENT

(PCT Administrative Instructions, Section 411)

From the INTERNATIONAL BUREAU

To:

IMAMURA, Masazumi  
KRF Building  
5th floor  
5-5, Kyobashi 1-chome  
Chuo-ku  
Tokyo 104-0031  
JAPON

Date of mailing (day/month/year) 23 November 1998 (23.11.98)			
Applicant's or agent's file reference 98215M		IMPORTANT NOTIFICATION	
International application No. PCT/JP98/04457		International filing date (day/month/year) 02 October 1998 (02.10.98)	
International publication date (day/month/year) Not yet published		Priority date (day/month/year) 02 October 1997 (02.10.97)	
Applicant ITAI, Akiko			
<p>1. The applicant is hereby notified of the date of receipt (except where the letters "NR" appear in the right-hand column) by the International Bureau of the priority document(s) relating to the earlier application(s) indicated below. Unless otherwise indicated by an asterisk appearing next to a date of receipt, or by the letters "NR", in the right-hand column, the priority document concerned was submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b).</p> <p>2. This updates and replaces any previously issued notification concerning submission or transmittal of priority documents.</p> <p>3. An asterisk(*) appearing next to a date of receipt, in the right-hand column, denotes a priority document submitted or transmitted to the International Bureau but not in compliance with Rule 17.1(a) or (b). In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.</p> <p>4. The letters "NR" appearing in the right-hand column denote a priority document which was not received by the International Bureau or which the applicant did not request the receiving Office to prepare and transmit to the International Bureau, as provided by Rule 17.1(a) or (b), respectively. In such a case, the attention of the applicant is directed to Rule 17.1(c) which provides that no designated Office may disregard the priority claim concerned before giving the applicant an opportunity, upon entry into the national phase, to furnish the priority document within a time limit which is reasonable under the circumstances.</p>			
<u>Priority date</u> 02 Octo 1997 (02.10.97)	<u>Priority application No.</u> 9/269611	<u>Country or regional Office or PCT receiving Office</u> JP	<u>Date of receipt of priority document</u> 20 Nove 1998 (20.11.98)

The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

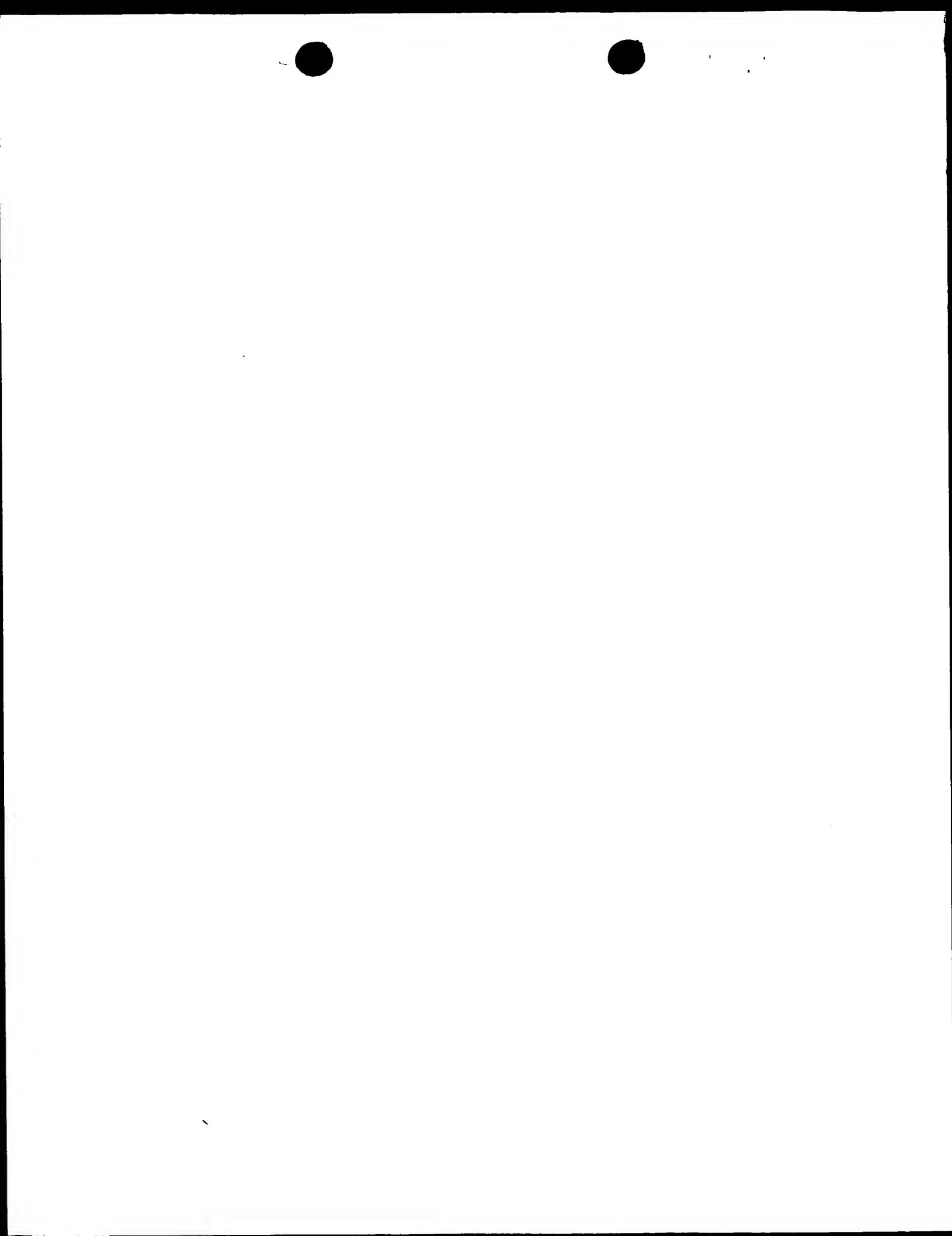
Facsimile No. (41-22) 740.14.35

Authorized officer

Sean Taylor

Telephone No. (41-22) 338.83.38





## PATENT COOPERATION TREATY

PCT

NOTICE INFORMING THE APPLICANT OF THE  
COMMUNICATION OF THE INTERNATIONAL  
APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

From the INTERNATIONAL BUREAU

To:

IMAMURA, Masazumi  
KRF Building  
5th floor  
5-5, Kyobashi 1-chome  
Chuo-ku  
Tokyo 104-0031 1999.4.23  
JAPON

Date of mailing (day/month/year)  
15 April 1999 (15.04.99)

Applicant's or agent's file reference  
98215M

## IMPORTANT NOTICE

International application No.  
PCT/JP98/04457

International filing date (day/month/year)  
02 October 1998 (02.10.98)

Priority date (day/month/year)  
02 October 1997 (02.10.97)

Applicant

ITAI, Akiko

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this Notice:

EP,JP,US

In accordance with Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:

None

The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this Notice is a copy of the international application as published by the International Bureau on  
15 April 1999 (15.04.99) under No. WO 99/18440

## REMINDER REGARDING CHAPTER II (Article 31(2)(a) and Rule 54.2)

If the applicant wishes to postpone entry into the national phase until 30 months (or later in some Offices) from the priority date, a **demand for international preliminary examination** must be filed with the competent International Preliminary Examining Authority before the expiration of 19 months from the priority date.

It is the applicant's sole responsibility to monitor the 19-month time limit.

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

## REMINDER REGARDING ENTRY INTO THE NATIONAL PHASE (Article 22 or 39(1))

If the applicant wishes to proceed with the international application in the **national phase**, he must, within 20 months or 30 months, or later in some Offices, perform the acts referred to therein before each designated or elected Office.

For further important information on the time limits and acts to be performed for entering the national phase, see the Annex to Form PCT/IB/301 (Notification of Receipt of Record Copy) and Volume II of the PCT Applicant's Guide.

The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer

J. Zahra

Telephone No. (41-22) 338.83.38



## PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

INFORMATION CONCERNING ELECTED  
OFFICES NOTIFIED OF THEIR ELECTION

(PCT Rule 61.3)

To:  
 IMAMURA, Masazumi  
 KRF Building  
 5th floor  
 5-5, Kyobashi 1-chome  
 Chuo-ku  
 Tokyo 104-0031 1999.4.23  
 JAPON

Date of mailing (day/month/year)  
 15 April 1999 (15.04.99)

Applicant's or agent's file reference  
 98215M

## IMPORTANT INFORMATION

International application No.	International filing date (day/month/year)	Priority date (day/month/year)
PCT/JP98/04457	02 October 1998 (02.10.98)	02 October 1997 (02.10.97)

Applicant  
 ITAI, Akiko

1. The applicant is hereby informed that the International Bureau has, according to Article 31(7), notified each of the following Offices of its election:

EP :AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE  
 National :JP,US

2. The following Offices have waived the requirement for the notification of their election; the notification will be sent to them by the International Bureau only upon their request:

None

3. The applicant is reminded that he must enter the "national phase" before the expiration of 30 months from the priority date before each of the Offices listed above. This must be done by paying the national fee(s) and furnishing, if prescribed, a translation of the international application (Article 39(1)(a)), as well as, where applicable, by furnishing a translation of any annexes of the international preliminary examination report (Article 36(3)(b) and Rule 74.1).

Some offices have fixed time limits expiring later than the above-mentioned time limit. For detailed information about the applicable time limits and the acts to be performed upon entry into the national phase before a particular Office, see Volume II of the PCT Applicant's Guide.

The entry into the European regional phase is postponed until 31 months from the priority date for all States designated for the purposes of obtaining a European patent including, where applicable, ES which cannot be elected since it is not bound by Chapter II.

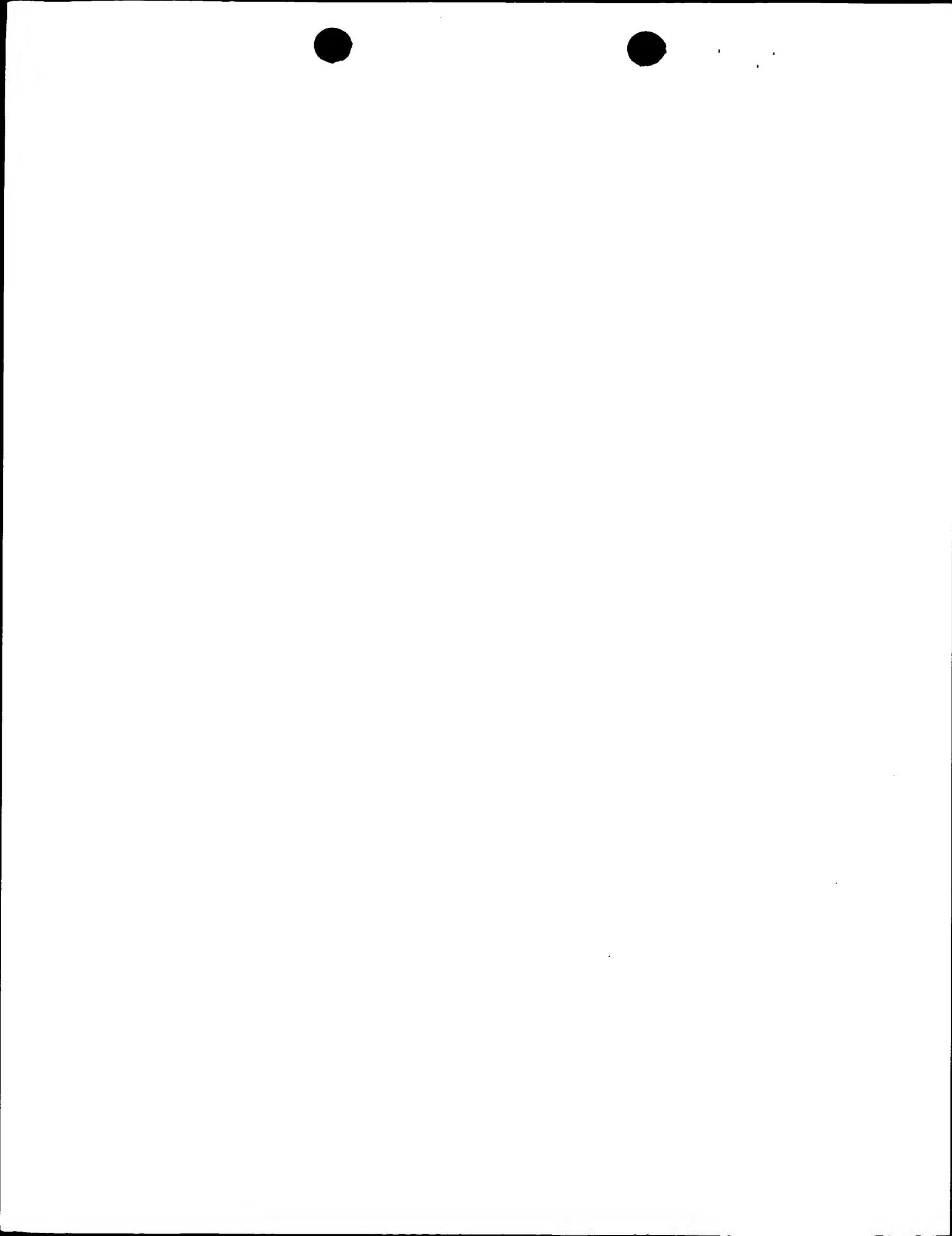
The International Bureau of WIPO  
 34, chemin des Colombettes  
 1211 Geneva 20, Switzerland

Facsimile No. (41-22) 740.14.35

Authorized officer:

J. Zahra

Telephone No. (41-22) 338.83.38



## 特許協力条約

発信人 日本国特許庁（国際予備審査機関）

出願人代理人

今村 正純

殿

あて名

〒 104-0031

東京都中央区京橋1丁目5番5号  
KRFビル5階

1998.1

SEARCHED, EXAMINED

SEARCHED

PCT見解書

(法第13条)  
(PCT規則66)発送日  
(日.月.年)

06.01.99

出願人又は代理人

の書類記号 98215M

応答期間

上記発送日から 2 月／母以内

国際出願番号

PCT/JP98/04457

国際出願日

(日.月.年)

02.10.98

優先日

(日.月.年)

02.10.97

国際特許分類 (IPC)

Int. C16 G01N33/68, G06F17/00, G06F15/18, C07K14/00

出願人 (氏名又は名称)

板井 昭子

1. これは、この国際予備審査機関が作成した 1 回目の見解書である。

2. この見解書は、次の内容を含む。

I  見解の基礎II  優先権III  新規性、進歩性又は産業上の利用可能性についての見解の不作成IV  発明の単一性の欠如V  法第13条 (PCT規則66.2(a)(ii)) に規定する新規性、進歩性又は産業上の利用可能性についての見解、それを裏付けるための文献及び説明VI  ある種の引用文献VII  国際出願の不備VIII  国際出願に対する意見

3. 出願人は、この見解書に応答することが求められる。

いつ?

上記応答期間を参照すること。この応答期間に間に合わないときは、出願人は、法第13条 (PCT規則66.2(d)) に規定するとおり、その期間の経過前に国際予備審査機関に期間延長を請求することができる。ただし、期間延長が認められるのは合理的な理由があり、かつスケジュールに余裕がある場合に限られることに注意されたい。

どのように? 法第13条 (PCT規則66.3) の規定に従い、答弁書及び必要な場合には、補正書を提出する。補正書の様式及び言語については、法施行規則第62条 (PCT規則66.8及び66.9) を参照すること。

なお 補正書を提出する追加の機会については、法施行規則第61条の2 (PCT規則66.4) を参照すること。

補正書及び/又は答弁書の審査官による考慮については、PCT規則66.4の2を参照すること。審査官との非公式の連絡については、PCT規則66.6を参照すること。

応答がないときは、国際予備審査報告は、この見解書に基づき作成される。

4. 国際予備審査報告作成の最終期限は、PCT規則69.2の規定により 02.02.00 である。

名称及びあて先 日本国特許庁 (IPEA/JP) 郵便番号 100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 亀田 宏之	2J	9015
	電話番号 03-3581-1101 内線 3252		



## 見解書

国際出願番号 PCT/JP98/04457

## I. 見解の基礎

1. この見解書は下記の出願書類に基づいて作成された。(法第6条(PCT14条)の規定に基づく命令に応答するために提出された差替え用紙は、この見解書において「出願時」とする。)

 出願時の国際出願書類

<input type="checkbox"/> 明細書 第 _____	ページ、	出願時に提出されたもの
明細書 第 _____	ページ、	国際予備審査の請求書と共に提出されたもの
明細書 第 _____	ページ、	付の書簡と共に提出されたもの
<input type="checkbox"/> 請求の範囲 第 _____	項、	出願時に提出されたもの
請求の範囲 第 _____	項、	PCT19条の規定に基づき補正されたもの
請求の範囲 第 _____	項、	国際予備審査の請求書と共に提出されたもの
請求の範囲 第 _____	項、	付の書簡と共に提出されたもの
<input type="checkbox"/> 図面 第 _____	ページ/図、	出願時に提出されたもの
図面 第 _____	ページ/図、	国際予備審査の請求書と共に提出されたもの
図面 第 _____	ページ/図、	付の書簡と共に提出されたもの
<input type="checkbox"/> 明細書の配列表の部分 第 _____	ページ、	出願時に提出されたもの
明細書の配列表の部分 第 _____	ページ、	国際予備審査の請求書と共に提出されたもの
明細書の配列表の部分 第 _____	ページ、	付の書簡と共に提出されたもの

2. 上記の出願書類の言語は、下記に示す場合を除くほか、この国際出願の言語である。

上記の書類は、下記の言語である \_\_\_\_\_ 語である。

- 国際調査のために提出されたPCT規則23.1(b)にいう翻訳文の言語
- PCT規則48.3(b)にいう国際公開の言語
- 国際予備審査のために提出されたPCT規則55.2または55.3にいう翻訳文の言語

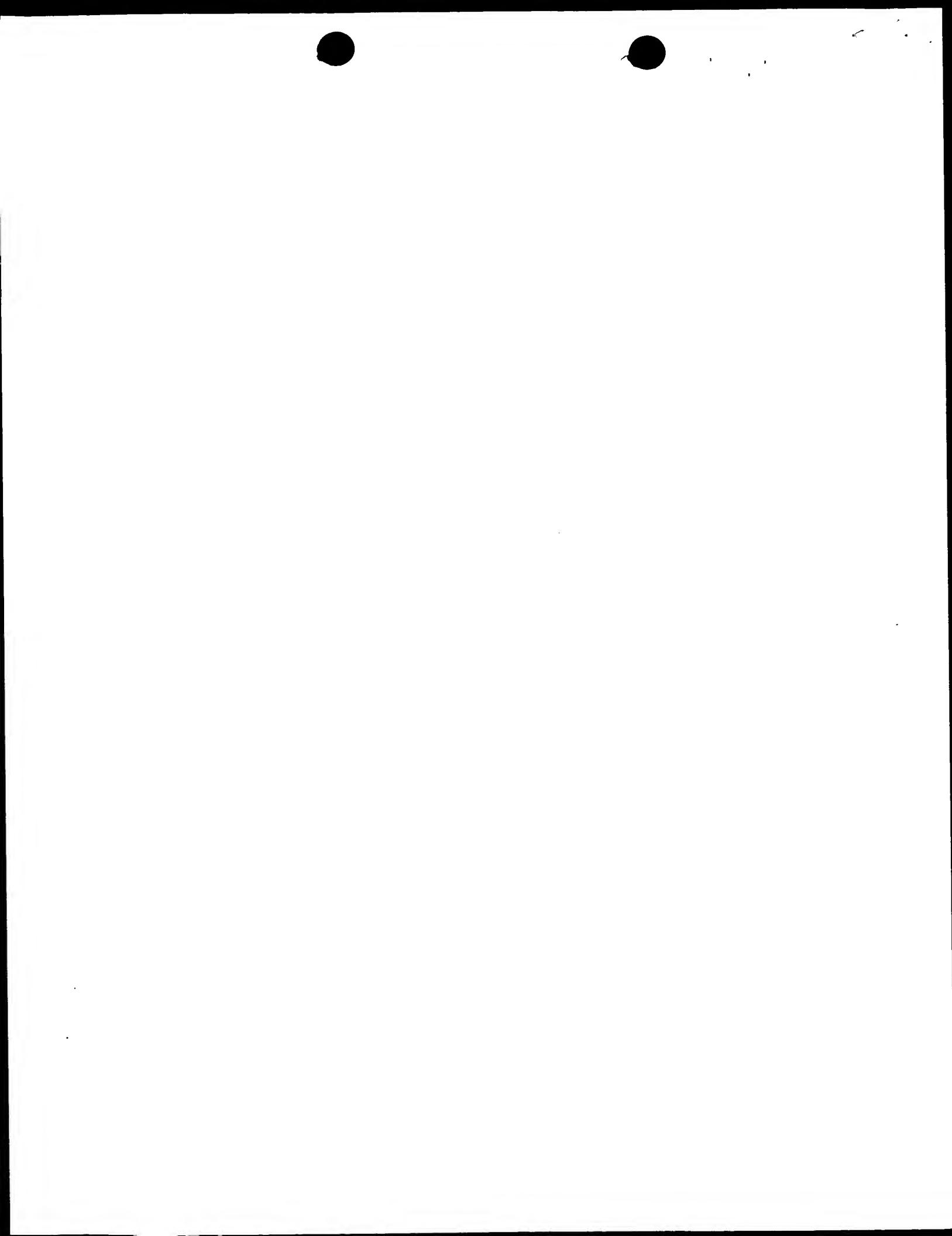
3. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき見解書を作成した。

- この国際出願に含まれる書面による配列表
- この国際出願と共に提出されたフレキシブルディスクによる配列表
- 出願後に、この国際予備審査（または調査）機関に提出された書面による配列表
- 出願後に、この国際予備審査（または調査）機関に提出されたフレキシブルディスクによる配列表
- 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった
- 書面による配列表に記載した配列とフレキシブルディスクによる配列表に記録した配列が同一である旨の陳述書の提出があった。

4. 補正により、下記の書類が削除された。

- 明細書 第 \_\_\_\_\_ ページ
- 請求の範囲 第 \_\_\_\_\_ 項
- 図面 図面の第 \_\_\_\_\_ ページ/図

5.  この見解書は、補充欄に示したように、補正が出願時における開示の範囲を越えてされたものと認められるので、その補正がされなかったものとして作成した。(PCT規則70.2(c))



## III. 新規性、進歩性又は産業上の利用可能性についての見解の不作成

1. 次に関して、当該請求の範囲に記載されている発明の新規性、進歩性又は産業上の利用可能性につき、次の理由により審査しない。

国際出願全体  
 請求の範囲 11

理由：

この国際出願又は請求の範囲 11 は、国際予備審査をすることを要しない次の事項を内容としている（具体的に記載すること）。

データベースは情報の単なる提示に相当する。

明細書、請求の範囲若しくは図面（次に示す部分）又は請求の範囲 の記載が、不明確であるため、見解を示すことができない（具体的に記載すること）。

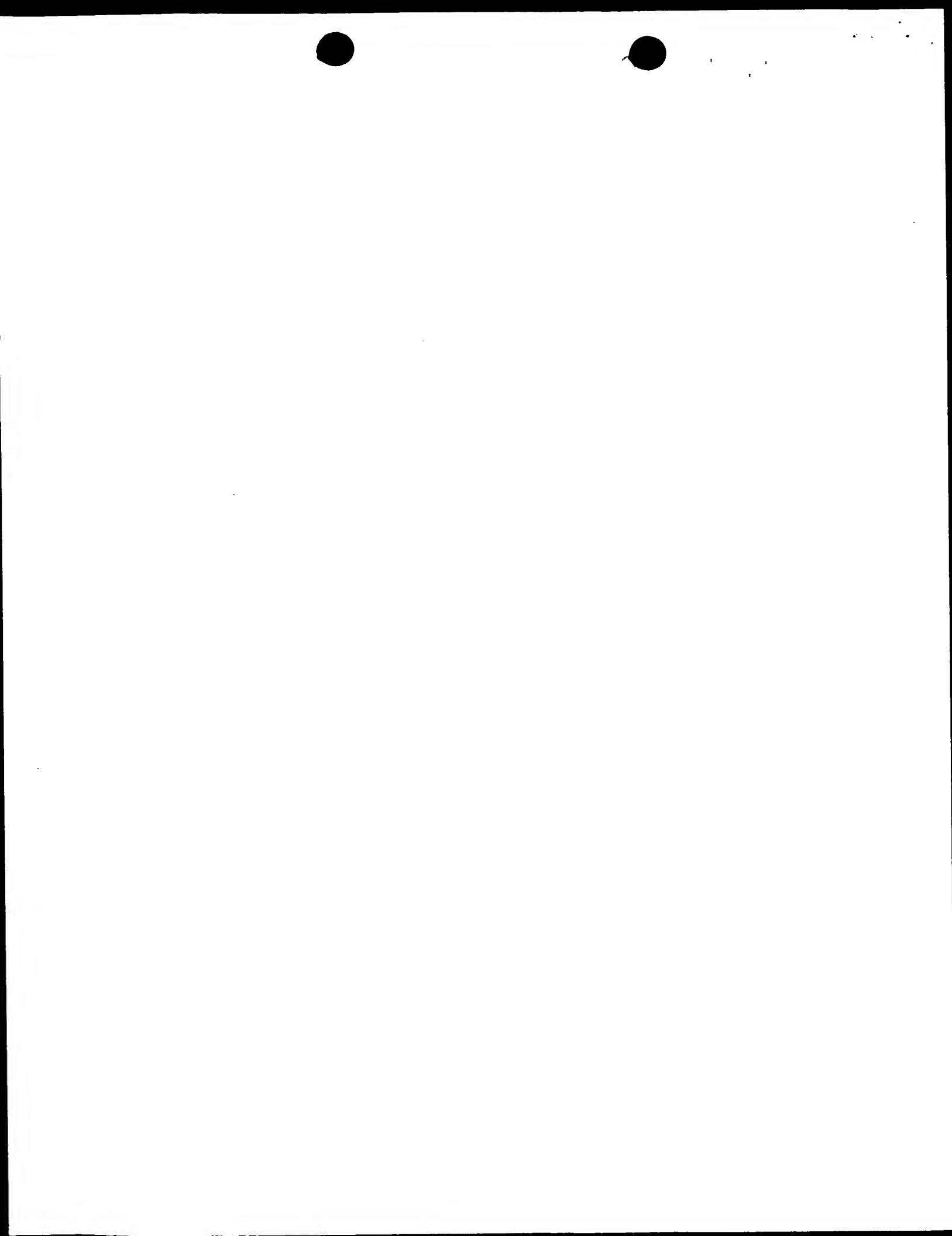
全部の請求の範囲又は請求の範囲 が、明細書による十分な裏付けを欠くため、見解を示すことができない。

請求の範囲 について、国際調査報告が作成されていない。

2. ヌクレオチド又はアミノ酸の配列表が実施細則の附属書C（塩基配列又はアミノ酸配列を含む明細書等の作成のためのガイドライン）に定める基準を満たしていないので、見解書を作成することができない。

書面による配列表が提出されていない又は所定の基準を満たしていない。

フレキシブルディスクによる配列表が提出されていない又は所定の基準を満たしていない。



## 見解書

国際出願番号 PCT/JP98/04457

V. 新規性、進歩性又は産業上の利用可能性についての法第13条 (PCT規則66.2(a)(ii)に定める見解、それを裏付る文献及び説明

## 1. 見解

新規性 (N)

請求の範囲 1~10 有  
請求の範囲 \_\_\_\_\_ 無

進歩性 (I S)

請求の範囲 1~10 有  
請求の範囲 \_\_\_\_\_ 無

産業上の利用可能性 (I A)

請求の範囲 1~10 有  
請求の範囲 \_\_\_\_\_ 無

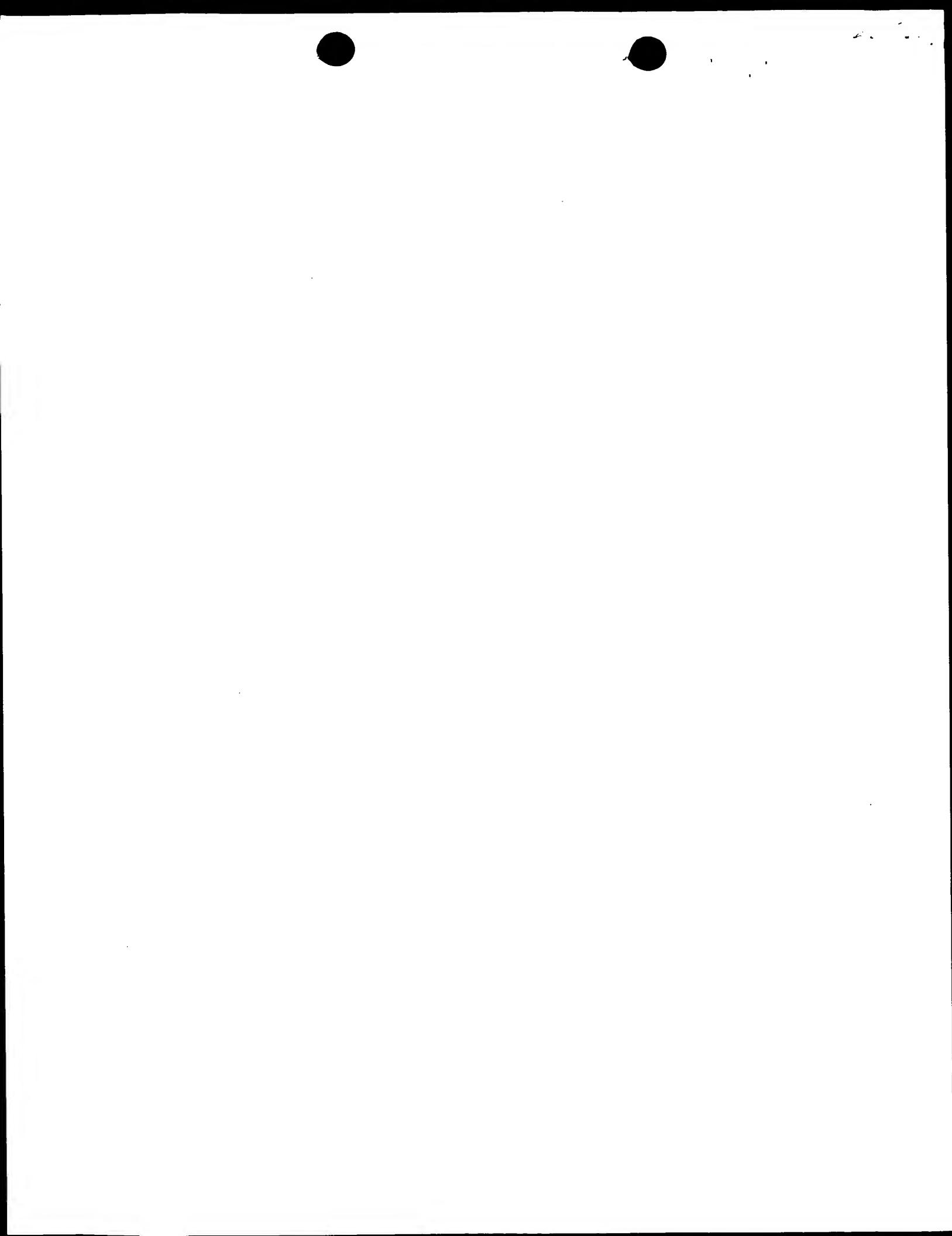
## 2. 文献及び説明

請求の範囲 1~10

文献1: J P, 7-206894, A (アドバンスドテクノロジーインスティチュート株式会社) 8. 8月. 1995 (08. 08. 95)

文献1には、蛋白質の立体構造の予測に際し、立体構造既知の蛋白質で相同性が高いものを選択して検討することが記載されている。その場合に、極性度やpK値等の物理化学的パラメータについて検討することも記載されている。

パラメータとしてアミノ酸残基の側鎖の疎水性や親水性を用いることは当業者にとって自明である。



特許協力条約に基づく出願出願

願書

出願人は、この国際出願が特許協力条約に従って処理されることを請求する。

国際出願番号	国際出願日
(受付印)	
出願人又は代理人の書類記号 (希望する場合、最大12字)	
98215M	

DCT  
02/10/98  
審査部

第Ⅰ欄 送りの名称

蛋白質の立体構造の推定方法

第Ⅱ欄 出願人

氏名(名称)及びあて名: (姓・名の順に記載: 法人は公式の完全な名称を記載; あて名は郵便番号及び国名も記載)

板井 昭子 ITAI Akiko

この欄に記載した者は、発明者でもある。

〒113-0033 日本国東京都文京区本郷5-16-6

電話番号:

5-16-6, Hongo, Bunkyo-ku, Tokyo

ファクシミリ番号:

113-0033 JAPAN

加入電信番号:

国籍(国名): 日本国 JAPAN

住所(国名): 日本国 JAPAN

この欄に記載した者は、次の  
指定国についての出願人である:

すべての指定国

米国を除くすべての指定国

米国のみ

追記欄に記載した指定国

第Ⅲ欄 その他の出願人又は送り明者

氏名(名称)及びあて名: (姓・名の順に記載: 法人は公式の完全な名称を記載; あて名は郵便番号及び国名も記載)

この欄に記載した者は  
次に該当する:

出願人のみである。

出願人及び発明者である。

発明者のみである。  
(ここにレ印を付したときは、以下に記入しないこと)

国籍(国名):

住所(国名):

この欄に記載した者は、次の  
指定国についての出願人である:

すべての指定国

米国を除くすべての指定国

米国のみ

追記欄に記載した指定国

その他の出願人又は発明者が続葉に記載されている。

第Ⅳ欄 代理人又は共通の代表者、通称のあて名

次に記載された者は、国際機関において出願人のために行動する:

代理人

共通の代表者

氏名(名称)及びあて名: (姓・名の順に記載: 法人は公式の完全な名称を記載; あて名は郵便番号及び国名も記載)

電話番号:

9621 弁理士 今村 正純 IMAMURA Masazumi

03-3271-1331

9263 弁理士 塩澤 寿夫 SHIOZAWA Hisao

ファクシミリ番号:

9584 弁理士 釜田 淳爾 KAMATA Junji

03-3271-1410

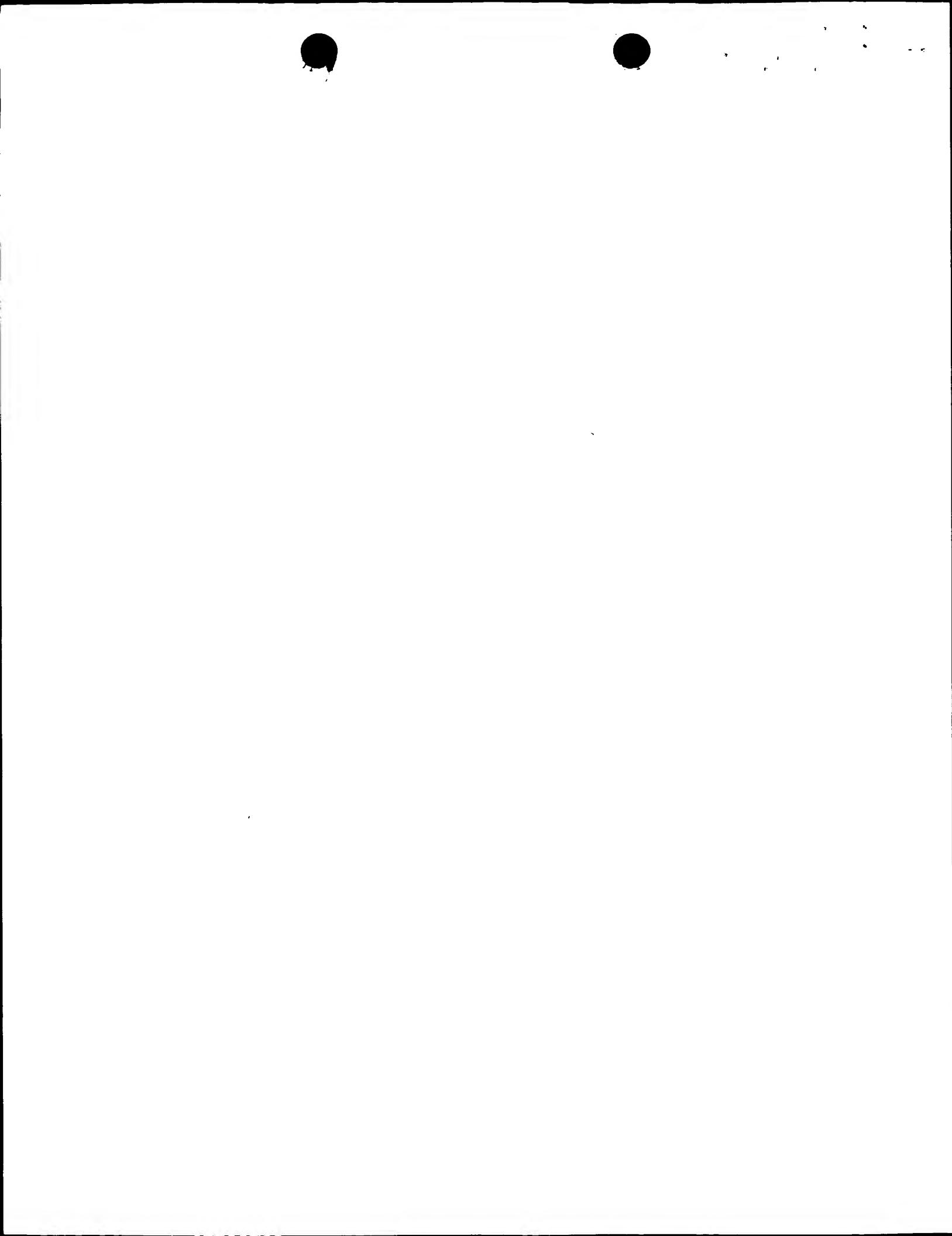
〒104-0031 日本国東京都中央区京橋1丁目5番5号

加入電信番号:

KRFビル5階

5th Floor, KRF Bldg., 5-5, Kyobashi 1-chome,  
Chuo-ku, Tokyo 104-0031 JAPAN

通知のためのあて名: 代理人又は共通の代表者が選任されておらず、上記欄内に特に通知が送付されるあて名を記載している場合は、レ印を付す



## 第二回　國の指揮

規則 4.9(4)の規定に基づき次の検定を行う(該当する口にレ印を付すこと: 少なくとも1つの口にレ印を付すこと)。

卷之三

**A P** **A R I P O** 4小国 : **G H** ガーナ Ghana, **G M** ガンビア Gambia, **K E** ケニア Kenya, **L S** レソト Lesotho, **M W** マラウイ Malawi, **S D** スーダン Sudan, **S Z** スワジランド Swaziland, **U G** ウガンダ Uganda, **Z W** ジンバブエ Zimbabwe, 及びハラレプロトコルと特許協力条約の締約国である他の国

**E A** ユーラシア半島 : **A M** アルメニア Armenia, **A Z** アゼルバイジャン Azerbaijan, **B Y** ベラルーシ Belarus, **K G** キルギス Kyrgyzstan, **K Z** カザフスタン Kazakhstan, **M D** モルドヴァ Republic of Moldova, **R U** ロシア Russian Federation, **T J** タジキスタン Tajikistan, **T M** トルクメニスタン Turkmenistan, 及びユーラシア特許条約と特許協力条約の締約国である他の国

**E P** ヨーロッパ半島 : **A T** オーストリア Austria, **B E** ベルギー Belgium, **C H** and **I I** スイス及びリヒテンシュタイン Switzerland and Liechtenstein, **C Y** キプロス Cyprus, **D E** ドイツ Germany, **D K** デンマーク Denmark, **E S** スペイン Spain, **F I** フィンランド Finland, **F R** フランス France, **G B** 英国 United Kingdom, **G R** ギリシャ Greece, **I E** アイルランド Ireland, **I T** イタリア Italy, **L U** ルクセンブルグ Luxembourg, **M C** モナコ Monaco, **N L** オランダ Netherlands, **P T** ポルトガル Portugal, **S E** スウェーデン Sweden, 及びヨーロッパ特許条約と特許協力条約の締約国である他の国

**O A** **O A P I** 半島 : **B F** ブルキナ・ファソ Burkina Faso, **B J** ベナン Benin, **C F** 中央アフリカ Central African Republic, **C G** コンゴ Congo, **C I** コートジボアール Côte d'Ivoire, **C M** カメルーン Cameroon, **G A** ガボン Gabon, **G N** ギニア Guinea, **M L** マリ Mali, **M R** モーリタニア Mauritania, **N E** ニジェール Niger, **S N** セネガル Senegal, **T D** チャド Chad, **T G** トーゴ Togo, 及びアフリカ知的所有権機構のメンバー国と特許協力条約の締約国である他の国 (他の層面の保護又は取扱いを求める場合は点線線上に記載する)

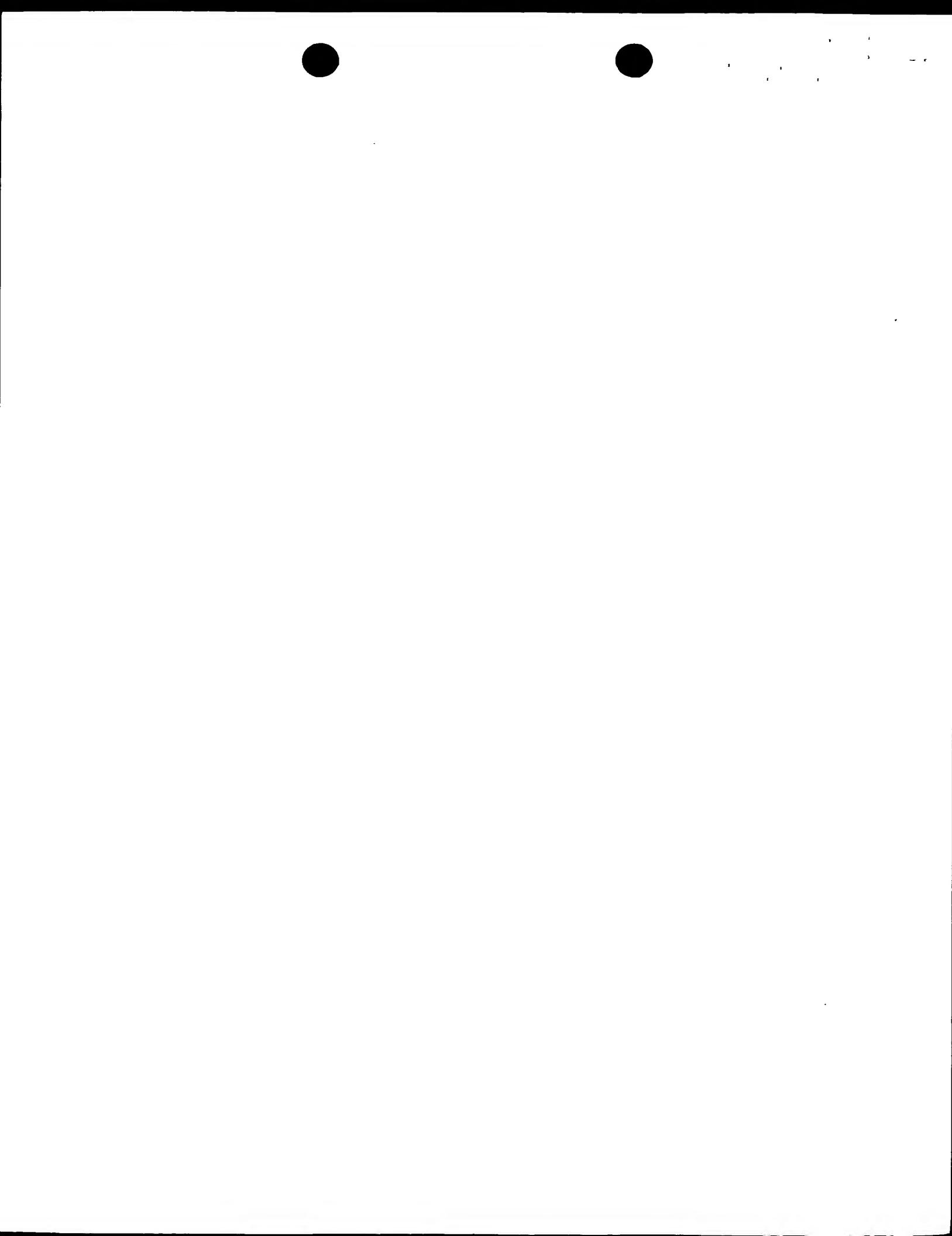
【圖1-6】空手道用（他の種類の保護とは戦扱いを含む場合には点線上に記載する）

<input type="checkbox"/> AL アルバニア Albania	<input type="checkbox"/> LT リトアニア Lithuania
<input type="checkbox"/> AM アルメニア Armenia	<input type="checkbox"/> LU ルクセンブルグ Luxembourg
<input type="checkbox"/> AT オーストリア Austria	<input type="checkbox"/> LV ラトヴィア Latvia
<input type="checkbox"/> AU オーストラリア Australia	<input type="checkbox"/> MD モルドバ Republic of Moldova
<input type="checkbox"/> AZ アゼルバイジャン Azerbaijan	<input type="checkbox"/> MG マダガスカル Madagascar
<input type="checkbox"/> BA ボスニア・ヘルツェゴヴィナ Bosnia and Herzegovina	<input type="checkbox"/> MK マケドニア旧ユーゴスラヴィア共和国 The former Yugoslav Republic of Macedonia
 -----	
<input type="checkbox"/> BB バルバドス Barbados	<input type="checkbox"/> MN モンゴル Mongolia
<input type="checkbox"/> BG ブルガリア Bulgaria	<input type="checkbox"/> MW マラウイ Malawi
<input type="checkbox"/> BR ブラジル Brazil	<input type="checkbox"/> MX メキシコ Mexico
<input type="checkbox"/> BY ベラルーシ Belarus	<input type="checkbox"/> NO ノルウェー Norway
<input type="checkbox"/> CA カナダ Canada	<input type="checkbox"/> NZ ニュージーランド New Zealand
<input type="checkbox"/> CH and LI スイス及びリヒテンシュタイン Switzerland and Liechtenstein	<input type="checkbox"/> PL ポーランド Poland
<input type="checkbox"/> CN 中国 China	<input type="checkbox"/> PT ポルトガル Portugal
<input type="checkbox"/> CU キューバ Cuba	<input type="checkbox"/> RO ルーマニア Romania
<input type="checkbox"/> CZ チェコ Czech Republic	<input type="checkbox"/> RU ロシア Russian Federation
<input type="checkbox"/> DE ドイツ Germany	<input type="checkbox"/> SD スーダン Sudan
<input type="checkbox"/> DK デンマーク Denmark	<input type="checkbox"/> SE スウェーデン Sweden
<input type="checkbox"/> EE エストニア Estonia	<input type="checkbox"/> SG シンガポール Singapore
<input type="checkbox"/> ES スペイン Spain	<input type="checkbox"/> SI スロヴェニア Slovenia
<input type="checkbox"/> FI フィンランド Finland	<input type="checkbox"/> SK スロヴァキア Slovakia
<input type="checkbox"/> GB 英国 United Kingdom	<input type="checkbox"/> SL シエラ・レオネ Sierra Leone
<input type="checkbox"/> GE グルジア Georgia	<input type="checkbox"/> TJ タジキスタン Tajikistan
<input type="checkbox"/> GH ガーナ Ghana	<input type="checkbox"/> TM トルクメニスタン Turkmenistan
<input type="checkbox"/> GM ガンビア Gambia	<input type="checkbox"/> TR トルコ Turkey
<input type="checkbox"/> GW ギニア・ビサオ Guinea-Bissau	<input type="checkbox"/> TT トリニダッド・トバゴ Trinidad and Tobago
<input type="checkbox"/> HR クロアチア Croatia	<input type="checkbox"/> UA ウクライナ Ukraine
<input type="checkbox"/> HU ハンガリー Hungary	<input type="checkbox"/> UG ウガンダ Uganda
<input type="checkbox"/> ID インドネシア Indonesia	<input type="checkbox"/> US 米国 United States of America
<input type="checkbox"/> IL イスラエル Israel	 -----
<input type="checkbox"/> IS アイスランド Iceland	<input type="checkbox"/> UZ ウズベキスタン Uzbekistan
<input type="checkbox"/> JP 日本 Japan	<input type="checkbox"/> VN ヴィエトナム Viet Nam
<input type="checkbox"/> KE ケニア Kenya	<input type="checkbox"/> YU ユーゴスラヴィア Yugoslavia
<input type="checkbox"/> KG キルギス Kyrgyzstan	<input type="checkbox"/> ZW ジンバブエ Zimbabwe
<input type="checkbox"/> KR 韓国 Republic of Korea	
<input type="checkbox"/> KZ カザフスタン Kazakhstan	
<input type="checkbox"/> LC セント・ルシア Saint Lucia	
<input type="checkbox"/> LK スリ・ランカ Sri Lanka	
<input type="checkbox"/> LR リベリア Liberia	
<input type="checkbox"/> LS レソト Lesotho	

以下の団は、この様式の施行後に特許協力条約の締約国となった国を指定（国内特許のために）するためのものである

<input type="checkbox"/>	.....

確認の指定の宣言：出席人は、上記の指定に加えて、規則 4. 9 (b) の規定に基づき、特許協力条約の下で認められる他の全ての国での指定を行う。ただし、この宣言から除く旨の表示を追記欄にした国は、指定から除外される。出席人は、これらの追加される指定が確認を条件としていること、並びに優先日から 1 月が経過する前にその確認がなされない場合は、この期間の経過時に、出席人によって取り下げられたものとみなされされることを宣言する。（指定の確認は、指定を特定する通知の提出と指定手数料及び特許手数料の提出からなる。この確認は、優先日から 1 月以内に受理官庁へ提出しなければならない。）



## 第VI欄 他の優先権の主張（先の出願）が追記欄に記載されている

先の出願日 (日、月、年)	先の出願番号	先の出願		
		国内出願：国名	広域出願：*広域官庁名	国際出願：受理官庁名
(1) 02. 10. 97	平成9年特許願 第269611号	日本国 JAPAN		
(2)				
(3)				

上記( )の番号の先の出願（ただし、本国際出願が提出される受理官庁に対して提出されたものに限る）のうち、次の( )の番号のものについては、出願書類の認証原本を作成し国際事務局へ送付することを、受理官庁（日本国特許庁の長官）に対して請求している。 (1)

\*先の出願が、ARIPOの特許出願である場合は、その先の出願を行った工業所有権の保護のためのパリ条約同盟国の中なくとも1ヶ国を追記欄に表示しなければならない（規則4. 10(b)(ii)）。追記欄を参照。

## 第VII欄 国際調査機関の機関名

国際調査機関の機関名 (ISA) の略記	当該調査機関の本邦語表記：当該調査の機関 (先の調査が国際調査機関によって既に実施又は請求されている場合)
ISA/JP	出願日 (日、月、年) 出願番号 国名 (又は広域官庁)

## 第VIII欄 国際出願：提出書類の箇目

この国際出願の用紙の枚数は次のとおりである。	
願書	3枚
明細書（配列表を除く）	18枚
請求の範囲	2枚
要約書	1枚
図面	1枚
明細書の配列表	0枚
合計	25枚

この国際出願には、以下にチェックした書類が添付されている。

1.  手数料計算用紙
5.  優先権書類（上記第VI欄の( )の番号を記載する）
2.  納付する手数料に相当する特許印紙を貼付した書面
6.  国際出願の翻訳文（翻訳に使用した言語名を記載する）
3.  国際事務局の口座への振込みを証明する書面
7.  寄託した微生物又は他の生物材料に関する書面
4.  別個の記名押印された委任状
8.  ネクレオチド又はアミノ酸配列表（フレキシブルディスク）
9.  記名押印（署名）の説明書

## 優先権書類送付請求書

要約書とともに提出する図面：

本国際出願の使用言語名：日本語

## 第IX欄 提出者の記名押印

各人の氏名（名称）を記載し、その次に押印する。

今村正純



塩澤寿夫



釜田淳爾



## 1. 国際出願として提出された書類の実際の受理の日

## 受理箇所の記入欄

## 3. 国際出願として提出された書類を補完する書類又は図面であって

その後期間内に提出されたものの実際の受理の日（訂正日）

## 4. 特許協力条約第11条(2)に基づく必要な補完の期間内の受理の日

## 5. 出願人により特定された

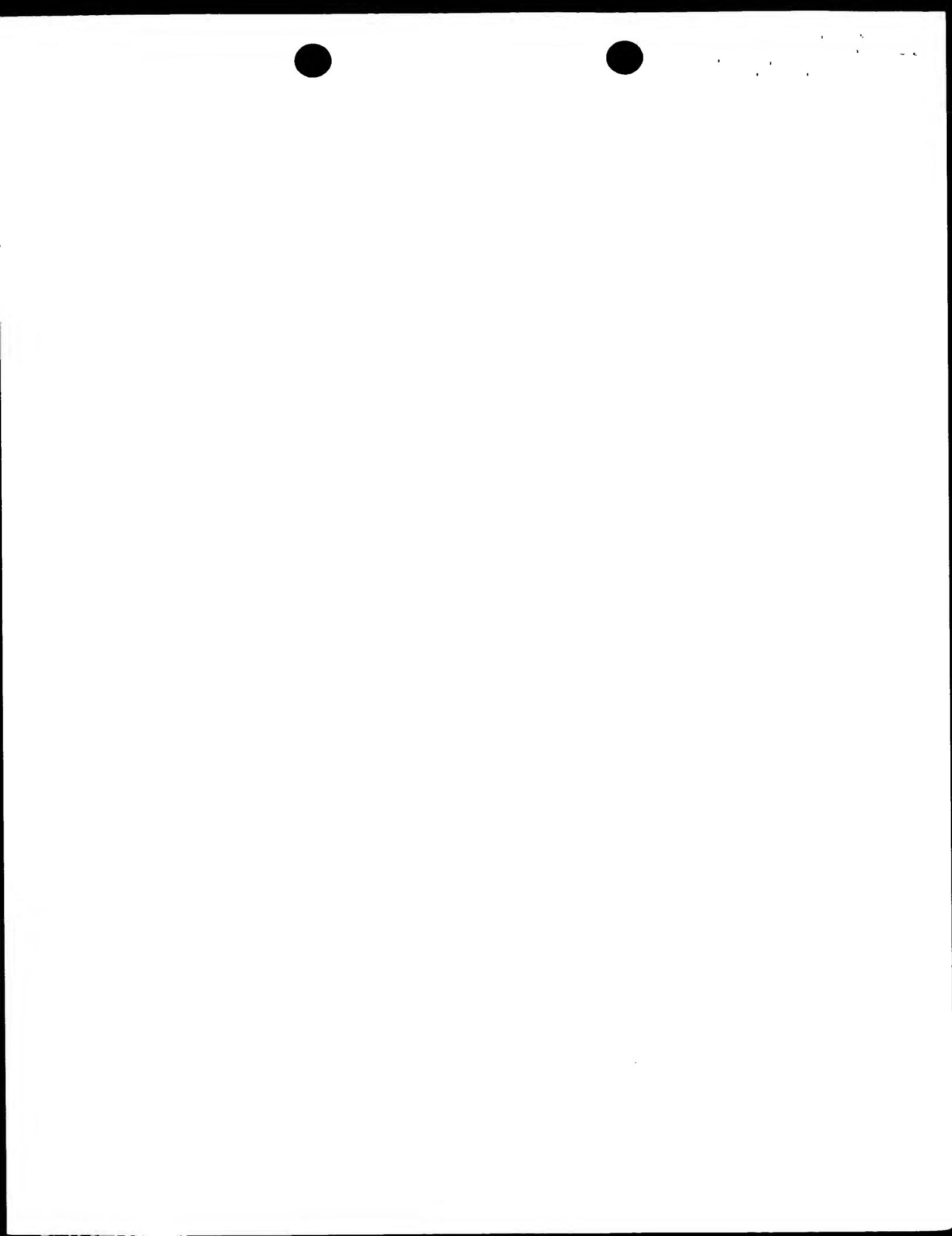
ISA/JP

6.  調査手数料未払いにつき、国際調査機関に調査用申しを送付していない

## 2. 図面

 受理された 不足図面がある

記録原本の受理の日	
株式会社PCT/RO/101 (最終用紙)	(1998年7月)



P C T

手 姓 本斗 昭子 氏名 月日 系氏  
顔 象 附 風 象

受理官庁記入欄

出願人又は代理人の書類記号  
98215M

国際出願番号
受理官庁の日付印

出願人  
板井 昭子

送付料の手数料の目次表

1. 及び 2. 特許協力条約に基づく国際出願等に関する法律(国内法)  
第18条第1項第1号の規定による手数料(注1)  
(送付手数料[T]及び調査手数料[S]の合計)

95,000 円 T+S

3. 国際手数料(注2)

基本手数料

国際出願に含まれる用紙の枚数 25 枚

最初の30枚まで ······

55,000 円 b1

0 × 1,300 =

0 円 b2

30枚を越える用紙の枚数 用紙1枚の手数料

55,000 円 B

b1及びb2に記入した金額を加算し、合計額をBに記入

指定手数料

国際出願に含まれる指定数(注3) 3

3 × 12,700 =

38,100 円 D

支払うべき指定手数料  
の数(上限は11)  
(注4)

93,100 円 I

D及びIに記入した金額を加算し、合計額をIに記入 ······

4. 納付すべき手数料の合計

T+S及びIに記入した金額を加算し、合計額を合計に記入

188,100 円

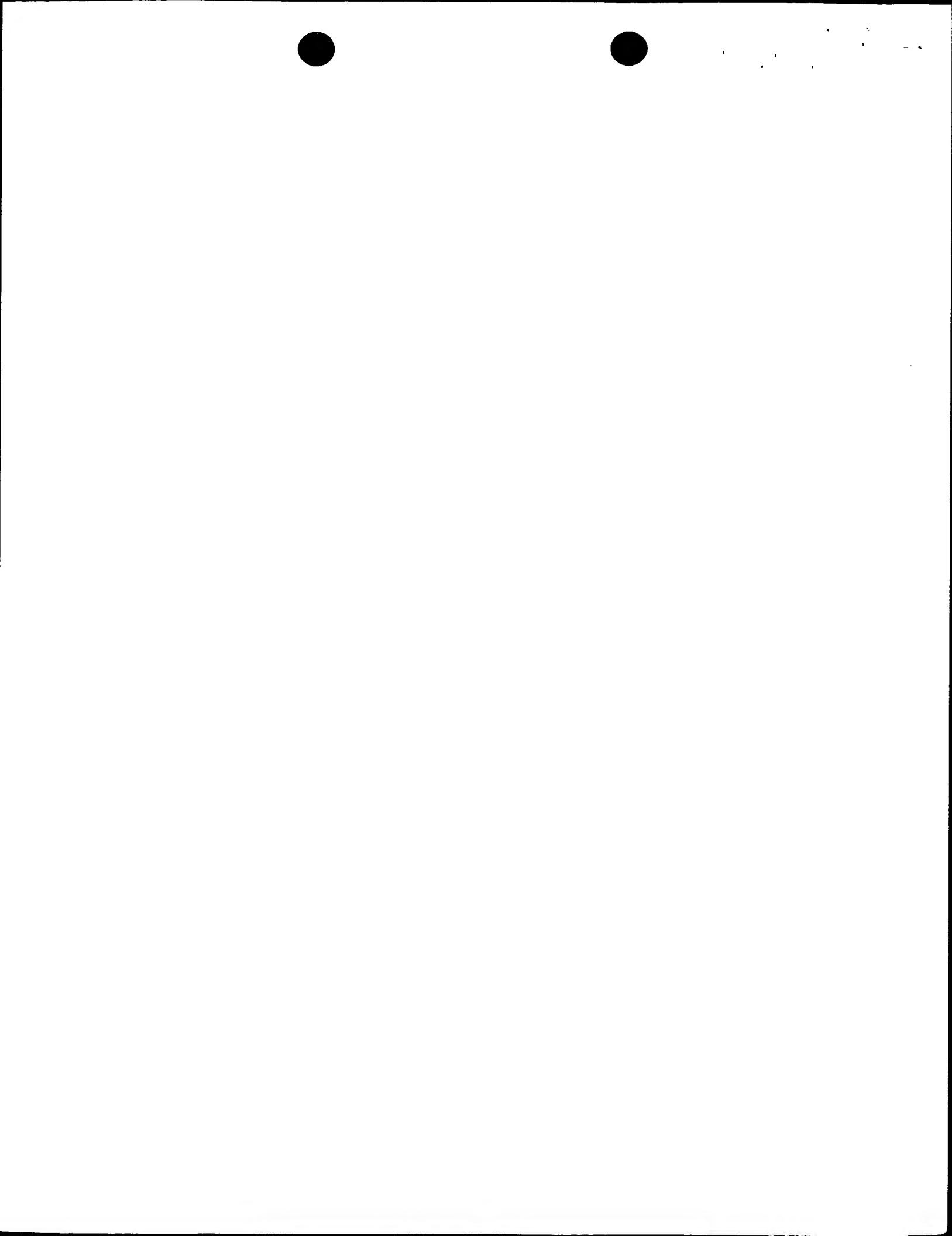
合 計

(注1) 送付手数料及び調査手数料については、合計金額を特許印紙をもって納付しなければならない。

(注2) 国際手数料については、受理官庁である日本国特許庁の長官が告示する国際事務局の口座への振込みを証明する書面を提出することにより納付しなければならない。

(注3) 顔書第V欄で印押を付した口の数。

(注4) 指定数を記入する。ただし、11指定以上は一律11とする。



## 明細書

### 蛋白質の立体構造の推定方法

#### 技術分野

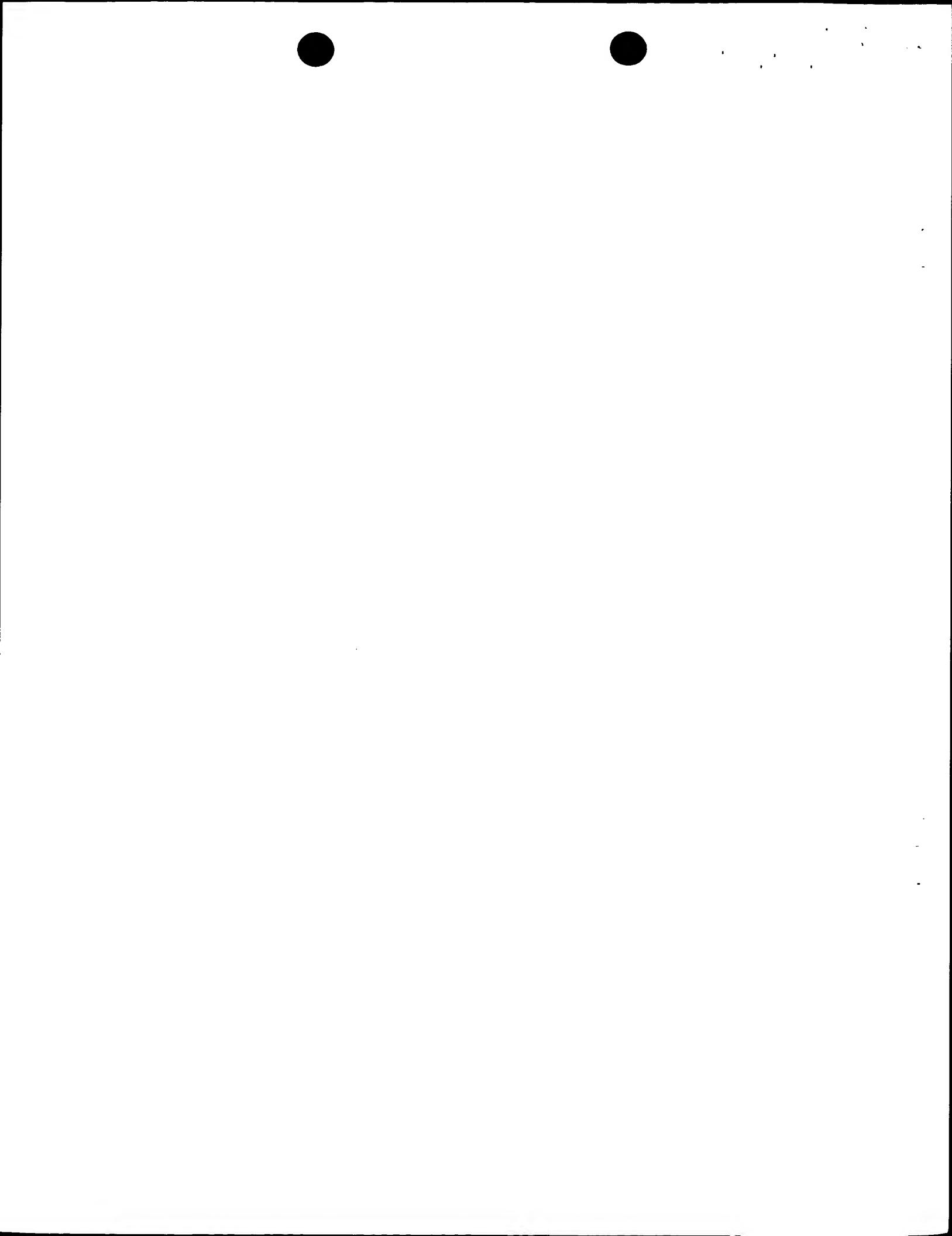
本発明は、蛋白質の立体構造の推定方法に関するものである。

#### 背景技術

アミノ酸配列から蛋白質の立体構造を推定することは理論的に不可能とはいえない。しかしながら、現在のところ、配列情報から蛋白質の立体構造を確実に推定する手段は開発されておらず、蛋白質の立体構造を知るための手段は、X線結晶構造解析や n m r 解析などの実験的方法に限られている。蛋白質の立体構造の情報は、その機能を原子レベルで理解し、またその蛋白質を標的とする医薬の創製やさらに優れた機能をもつ有用な蛋白質の創製に不可欠である。近年、遺伝子情報の解析手段が急速に進歩した結果、実際に蛋白質が単離されないまま配列情報だけが解明される例が急増している。従って、配列情報から蛋白質の立体構造や機能を推定する有効な手段の開発が切望されているのが現状である。

あるアミノ酸配列を有する蛋白質の存在がわかった場合、配列データベースから相同性のある蛋白質を検索するのが一般的である。アミノ酸配列の一致度がある程度よい蛋白質が見つかると、さらにその蛋白質との間で相同性やギャップも考慮したアラインメントを行い、さらに相同性の高いアラインメントの探索が行われる。目的の蛋白質と機能既知の蛋白質との相同性が高い場合には機能がその既知蛋白質に類似しており、一方、立体構造既知の蛋白質との相同性が高い場合には立体構造がその蛋白質に類似しているとの推定が成り立つ。また、相同性が高ければ高いほど機能や立体構造が類似している確率が高く、その推定の信頼性が高いと考えられる。

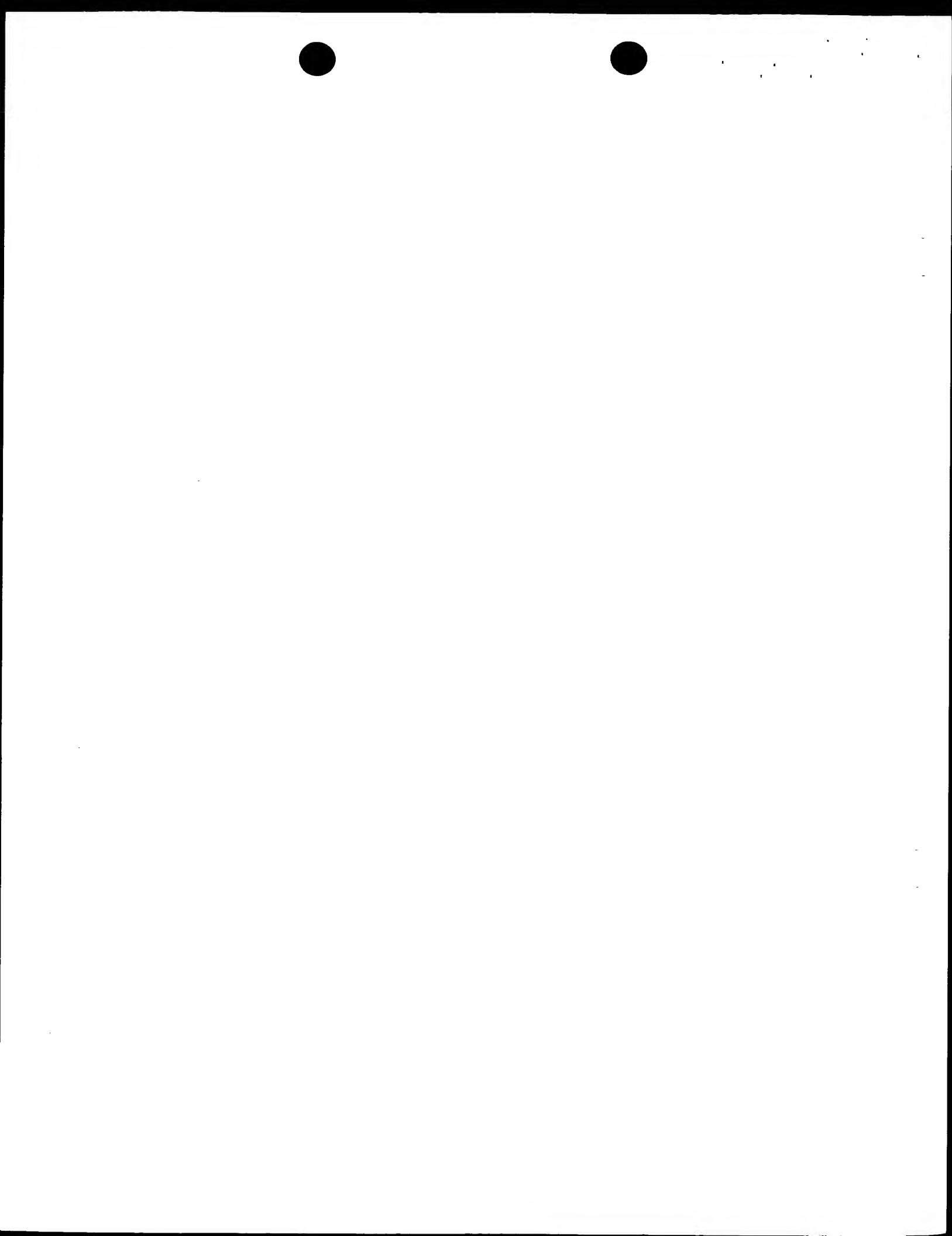
立体構造既知の蛋白質の配列とある程度（一般に 30% 程度）以上の相同性が



認められた場合には、その立体構造を鋳型として立体構造を構築するホモロジーモデリング法が行われる。鋳型の立体構造に基づいて対応付けられた残基が鋳型と異なる場合には、側鎖を置換することによって仮想的な立体構造を構築することができる。アラインメント上のギャップは、鋳型立体構造中に対応するアミノ酸がないか、あるいは鋳型の方が余計なアミノ酸をもつことになるので別途処理する必要があるが、ギャップの存在は鋳型に基づくモデリング作業を難しくし、かつ信頼性を損なうので、できるだけギャップの数を少なくするようにギャップにペナルティを与えるアラインメント法が推奨されている。

問題のアミノ酸配列とある程度以上相同性の高いアミノ酸配列を有する立体構造既知の蛋白質が見つからない場合には、ホモジーモデリングは不可能である。一方で、蛋白質の結晶構造情報が蓄積されるにつれて、アミノ酸配列の相同性が殆どなく機能も全く異なる複数の蛋白質が、類似の立体構造を持つ例が多数わかつてき。このことは、蛋白質が安定な立体構造をとるための物理化学的要因を考慮すれば、アミノ酸配列の相同性が低い場合にも、立体構造既知の蛋白質群中から鋳型となる立体構造を見出せる可能性があることを示している。

近年、アミノ酸残基毎の疎水性など物理的な性質の一致を考慮したスコアを用いることにより、アミノ酸配列の相同性が低くても立体構造的に類似性が高いと推定される鋳型蛋白質を立体構造既知の蛋白質群から選ぶ方法が開発された。代表的な方法として、Eisenberg らによる 3D-1D 法 (R. Luthy, J. U. Bowie and D. Eisenberg, *Nature*, 356, 83, 1992) がある。この方法は、アミノ酸配列の相同性に加え、立体構造既知の蛋白質について各アミノ酸残基の属する 2 次構造とその残基位置の環境を表すパラメータと、問題の配列の各アミノ酸残基に対して 2 次構造毎に与えたパラメータとを用いて、対応づけられたアミノ酸残基間で類似性のスコアを計算する工程を含んでいる。この方法では、蛋白質のペプチド鎖の折り畳み方についての膨大な自由度の問題を既知の結晶構造群を鋳型として用いることで回避しており、評価の要素として疎水性などの物理的パラメータを含めることによって配列の相同性が低い場合にもモデリングが可能である。



しかしながら、立体構造が類似している場合であっても、アミノ酸残基数、2次構造、及びループの長さが同じことは少ないので、アミノ酸配列間の対応づけに基づいた3D-1D法を実際に適用する場合には数多くの問題が生じる。例えば、アミノ酸配列間の単純なスライド（スレッディング）に加え、どちらかの配列に部分的な配列の欠損など（ギャップ）を考慮して対応付けする必要があるが、ホモロジーモデリングと同様にギャップの導入はモデリングの信頼性を低下させるてしまう。配列の相同性が低いときに、必要にして最小限のギャップを考慮した対応付けをいかにして実現するかが問題である。また、疎水性親水性などのパラメータの他、20種類のアミノ酸残基に対して2次構造毎のパラメータを用いるなど非常に多数のパラメータに依存しており、パラメータの改良による予測性の向上は期待しにくい。

アミノ酸配列から蛋白質の立体構造を予測する研究の歴史は、配列のどの部分がどのような2次構造をとるかを予測することから始まった。多数の蛋白質の結晶構造情報から統計的に求められたアミノ酸残基毎あるいは連続した数アミノ酸残基の組毎に、 $\alpha$ -ヘリックス又は $\beta$ -シートになり易さを示すパラメータを用いて、問題のアミノ酸配列から顕著な傾向を示す連続した領域を検出し、それがどちらの構造をとるかを予測するものである。その代表的なものとして、ChouとFasmanらによる2次構造予測法（P. Y. Chou, & G. D. Fasman, *Adv. enzymol.* 47, 45, 1978）がある。しかしながら、このような方法は、2次構造の3次元的集合に関してはなんら情報を与えるものではなく、また予測された2次構造と結晶解析で確認された2次構造の一致が60%前後であることから、立体構造の推定方法としてはほとんど利用価値がない。

蛋白質の安定な折り畳み構造を、純粋に計算的手法によって先入観を入れずに予測する試みも行われるようになった（いわゆる *ab initio* 予測法）。しかしながら、蛋白質は極めて自由度が大きい分子であり（100残基程度の蛋白質でも考慮すべき自由度のパラメータは400以上である）、すべての自由度を考慮して可能な構造を十分に探索することは、現在利用可能なコンピュータでは不可能で



ある。また、可能な立体構造の安定性を正しく評価できるほど蛋白質の構造の安定化に関わるファクター（例えば、水の物理化学的性質、疎水相互作用、静電相互作用）についての研究は進んでいないなどの理由から、このような構造予測法の成功は今のところ期待できない。

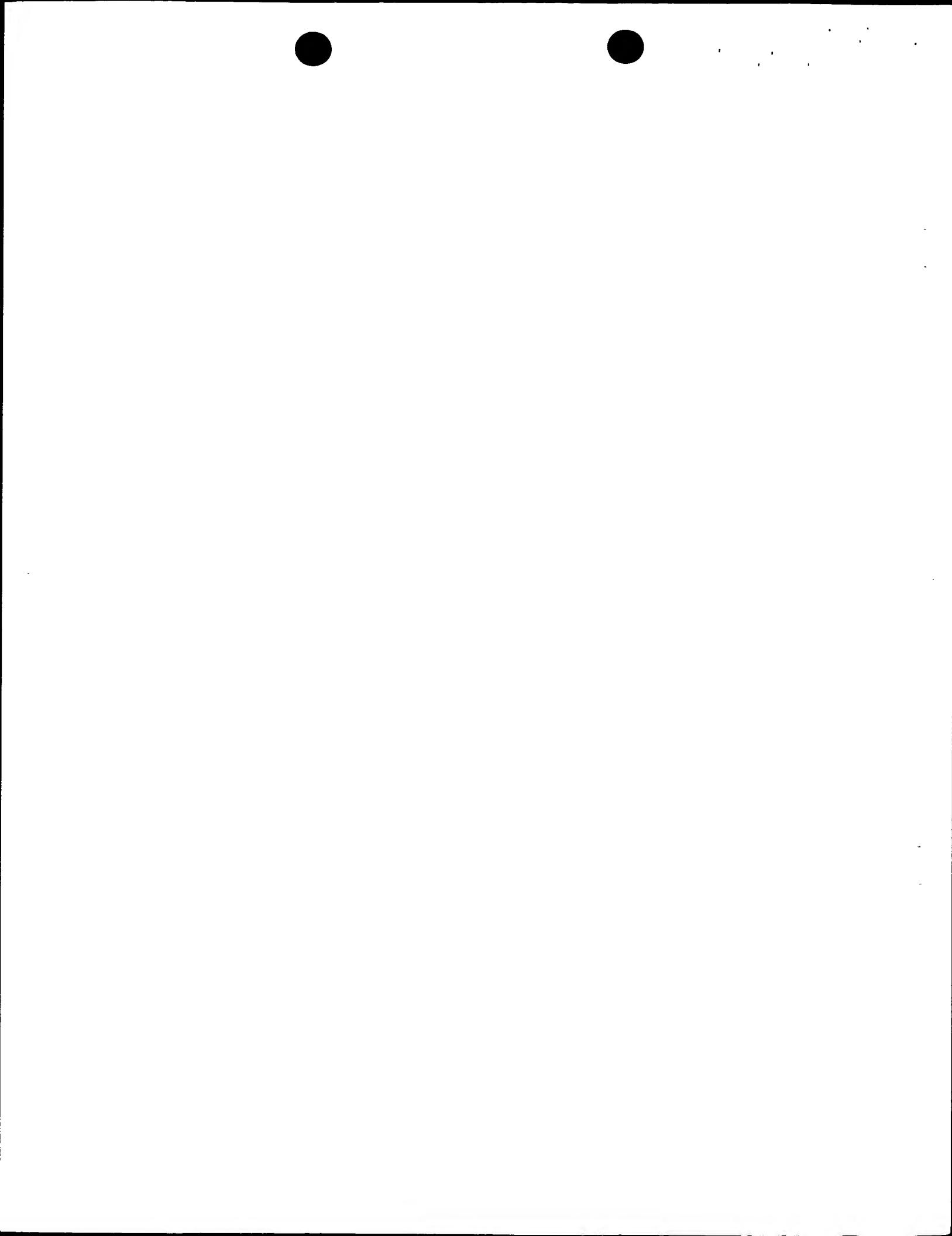
一方、近年、多数の蛋白質の立体構造が解析されており、その成果はプロテインデータバンクから利用できるようになっている。現在約 6,000 の蛋白質・核酸の構造が収録されており、機能の異なる独立の蛋白質は 400 程度である。これらの蛋白質の立体構造から、配列の相同性もなく進化的にも機能的にも全く関係がないように見える蛋白質が同じ構造モチーフを有する例が多数明らかにされている。

#### 発明の開示

本発明の第一の課題は、あるアミノ酸配列を有する蛋白質のアミノ酸配列情報から、その蛋白質がとる可能性の高いスキヤッフォールドを推定することにより、立体構造をモデリングするための方法を提供することにある。正しいスキヤッフォールドの推定は、立体構造を正しく必要な精度でモデリングするための出発点となり得る。すなわち、本発明の究極の課題は、アミノ酸配列の情報のみから蛋白質の立体構造を高い信頼性をもって推定する方法を提供することにあり、その手段としてスキヤッフォールドを推定する方法を提供することが本発明の具体的課題である。また、本発明の別の課題は、上記の方法に有用なデータベースを提供することにある。

本発明者らは上記の課題を解決すべく銳意努力した結果、立体構造が既知または推定可能な蛋白質について各アミノ酸残基の側鎖の環境情報を備えたデータベースを作成し、そのデータベースを利用することによって、立体構造未知の蛋白質のアミノ酸配列の情報から、その蛋白質がとる可能性の高いスキヤッフォールドを信頼性高く効率的に推定することができる方法を見出した。

すなわち本発明は、立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列



に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキヤップフォールドを推定する方法を提供するものである。スキヤップフォールドを推定した後に、鋳型蛋白質と質問配列の最適マッチングに基づいて質問配列に対応した立体構造（3次元座標）を得ることができる。

本発明の好ましい態様によれば、参照蛋白質のアミノ酸配列が該参照蛋白質の立体構造に基づいて連続した2以上のアミノ酸残基からなる2以上の部分配列に分割された上記方法；参照蛋白質のアミノ酸配列が疎水コアの形成に実質的に関与する1又は2以上のコア部分配列と疎水コアの形成に実質的に関与しない1又は2以上のサブ部分配列とに分割された上記方法；参照蛋白質における各アミノ酸残基の側鎖の蛋白質内部への埋没度及び／又は蛋白質表面への露出度の情報と、質問配列の各アミノ酸の疎水性及び／又は親水性の性質とに基づいてマッチングを行う上記方法；参照蛋白質の1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の片端又は両端以外においてはギャップを考慮せずにマッチングを行う上記方法；並びに、ギャップが1又は2以上のアミノ酸残基の削除又は付加である上記方法が提供される。

本発明のさらに好ましい態様によれば、マッチングが以下の工程：

- (a) 1又は2以上のコア部分配列を質問配列上でスライドさせながら、必要に応じて該コア部分配列の片端又は両端においてはギャップを考慮してマッチングを行う工程（ただし、2以上のコア部分配列を用いる場合には、該コア部分配列は参照蛋白質のアミノ酸配列での出現順に順番に配置する）；及び
- (b) 工程(a)でコア部分配列にマッチングされなかった質問配列の部分につき、該参照配列における各コア部分配列とのつながりにしたがって、必要に応じてギャップを考慮しつつサブ部分配列を質問配列上でスライドさせる工程を含む上記方法；参照蛋白質のアミノ酸残基の側鎖に関する環境情報と質問配列上の対応ア



ミノ酸残基の疎水度パラメータとから計算したスコアに基づいて最適マッチングを選択する上記方法；並びに、参照蛋白質についての自己マッチングスコアを用いて上記スコアを規格化する工程を含む上記方法が提供される。

別の観点からは、本発明により、立体構造が既知又は推定可能な1又は2以上の参照蛋白質のアミノ酸残基の側鎖についての環境情報を含むデータベースであって上記の各方法に用いられるためのデータベースが提供される。このデータベースは、通常の記憶媒体、例えば磁気ディスク、光ディスク、CD-ROM、磁気テープなどに格納された形態で流通可能であり、該環境情報と質問配列のアミノ酸残基の性質との一致度をスコアとするマッチングによって、質問配列から構成される蛋白質の安定なスキヤッフォールドを推定するために用いることができる。

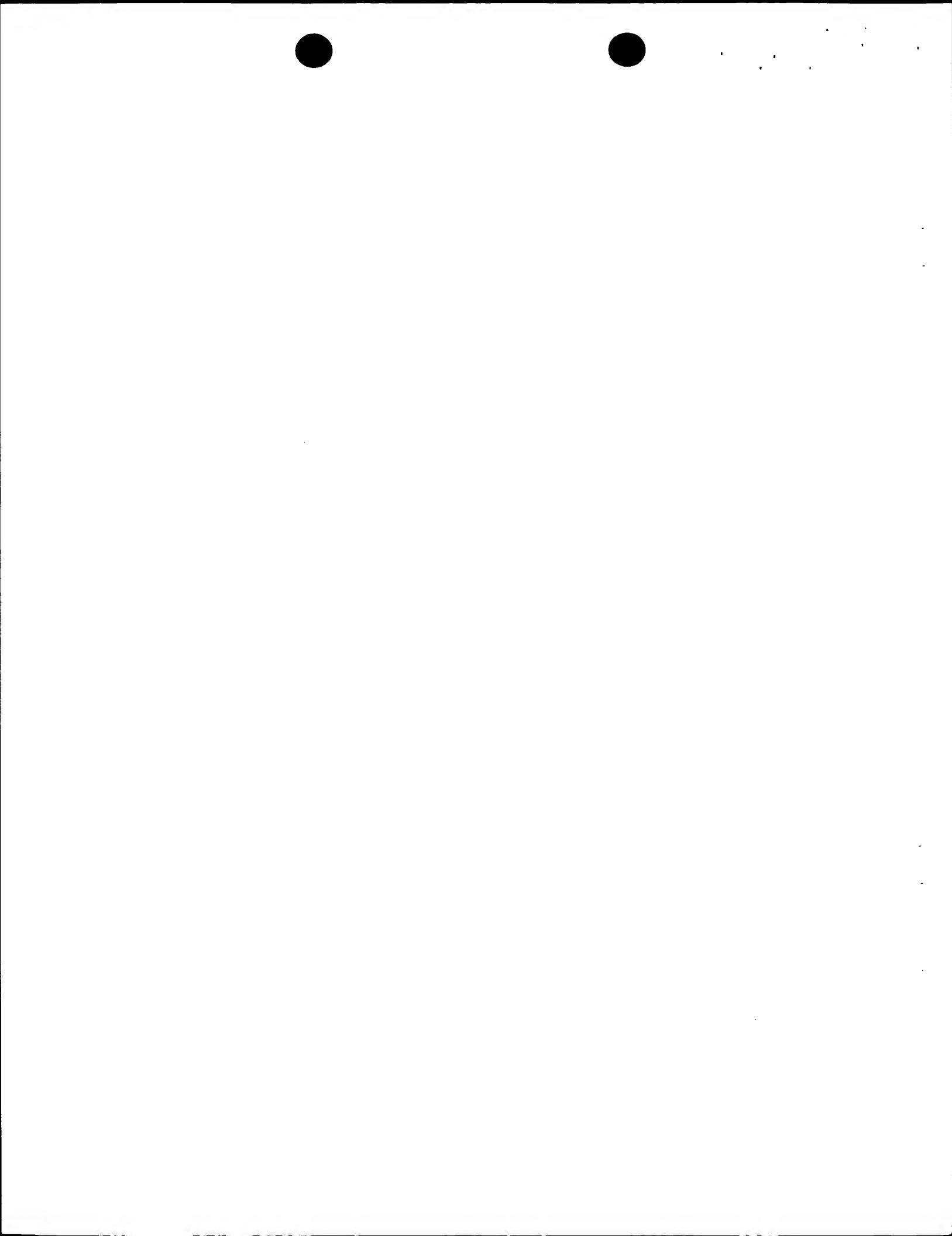
#### 図面の簡単な説明

第1図は、「立体構造」、「構造モチーフ」、及び「スキヤッフォールド」の関係を示した図である。図中、(a) は立体構造を示し、(b) は構造モチーフを示し、(c) はスキヤッフォールドを示す。

#### 発明を実施するための最良の形態

本発明の方法は、質問配列から理論的に可能な主鎖の折り畳み方を網羅・探索して評価するかわりに、参照蛋白質のデータベースから質問配列の蛋白質が立体構造的に類似する鑄型蛋白質を選び、鑄型蛋白質のスキヤッフォールドから質問配列の蛋白質の立体構造を推定することを特徴としている。本発明の方法は、一般的には、ワークステーション、パーソナルコンピューターなどの汎用のコンピューターを用いて高速に行うことができる。

本明細書において用いられる用語は、以下に述べる概念を含めて、最も広義に解釈する必要がある。「立体構造」とは3次元座標で表される蛋白質の構造を意味しており、アミノ酸残基の側鎖など存在する全原子を含む場合もあるが、それらの一部を省略することもある。「モデリング」とは、立体構造が実験的に解明

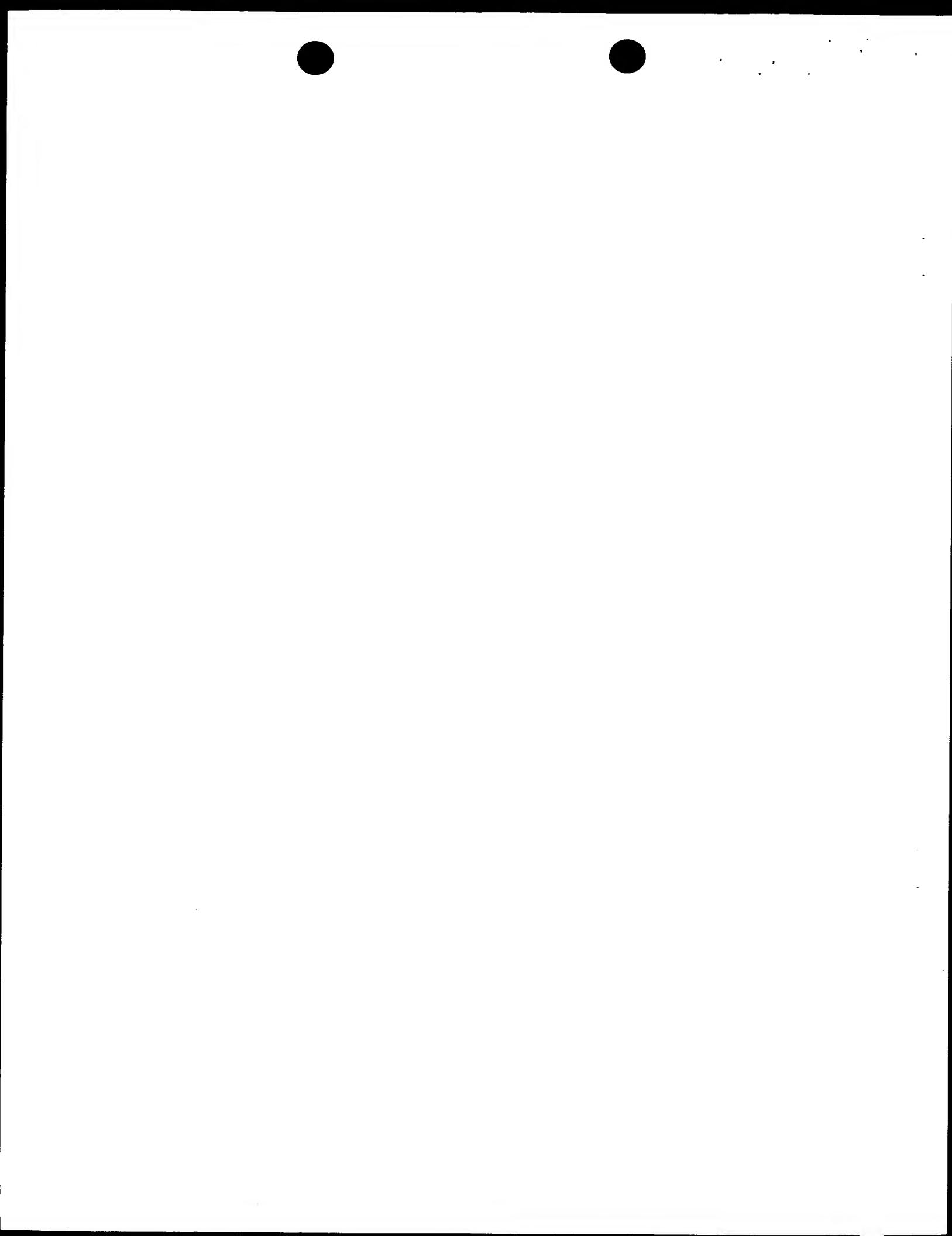


されているか否かに係わらず、ある蛋白質について存在可能性の高い立体構造を構築して原子の3次元座標で表現することをいう。

「2次構造」、「構造モチーフ」、及び「スキヤッフォールド」などの用語は Branden 及び Tooze らの著書に記載されている (Carl Branden and John Tooze, *Introduction to Protein Structure*, Garland publishing Inc. New York, 1991; 和訳「タンパク質の構造入門」、教育社、1992)。「構造モチーフ」と「スキヤッフォールド」は、ともにペプチド主鎖のみのトポロジーを表す点では共通しているが、「構造モチーフ」が2次構造群の組み合わせとペプチド鎖の流れを平面的かつ模式的に表現するのに対して、「スキヤッフォールド」は2次構造群の集合状態を含む蛋白質の3次元構造の骨組を意味する。「立体構造」、「構造モチーフ」、及び「スキヤッフォールド」の関係を図1に示す。

立体構造を推定したいアミノ酸配列を「質問配列」と呼び、その質問配列から構成される蛋白質を「質問配列の蛋白質」と呼ぶ。立体構造が既知または推定可能な蛋白質であって本発明のデータベースに含まれる蛋白質を「参照蛋白質」と呼び、参照蛋白質のうち、質問配列とのマッチングスコアがよく、質問配列の蛋白質と立体構造の類似性が高い蛋白質として選別された参照蛋白質を「鋳型蛋白質」という（「鋳型候補蛋白質」と呼ぶ場合もある）。鋳型蛋白質のスキヤッフォールドは、質問配列の蛋白質の立体構造を構築する際の鋳型として用いられる。

一般に2以上の配列をアミノ酸残基の一致度や相同性が高くなるように対応づけて並置する操作又は結果は「アラインメント」と呼ばれるが（「一致度」とは配列間で対応する残基間の厳密な一致を意味しており、「相同性」とは同等または類似など曖昧さを含めた一致の割合を意味する）、本発明の方法において「マッチング」（又は「対応付け」）という場合には、アミノ酸残基の一致又は相同性によらず、環境情報と性質の一致で残基を対応づけて並置する操作又は結果を意味している。ある対応付けにおける参照蛋白質の各アミノ酸残基の環境と質問配列のアミノ酸残基の性質との一致度を「マッチングスコア」（又は単に「スコア」）とよぶ。「環境情報」は、主として、参照蛋白質の立体構造における各アミノ酸

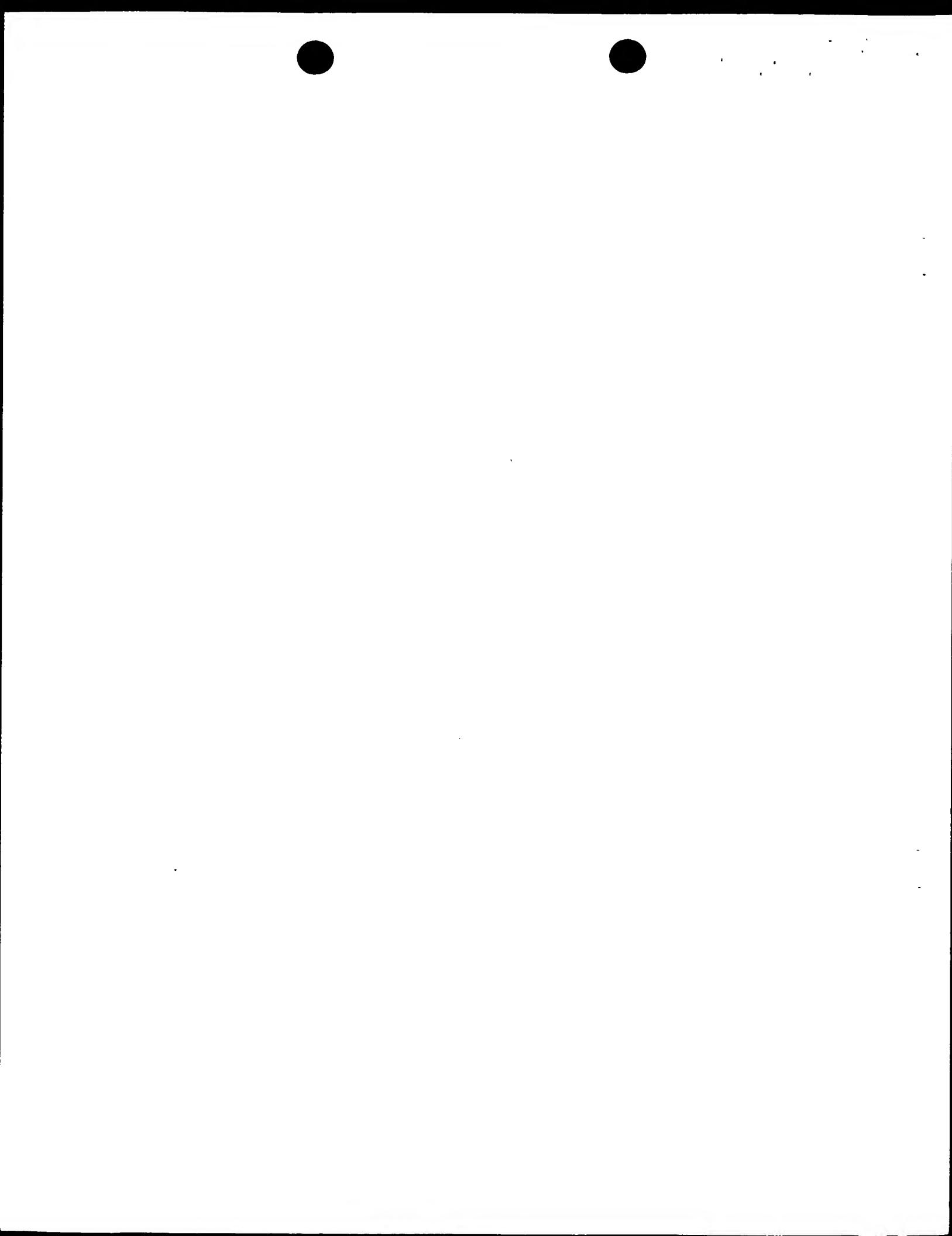


残基の側鎖の蛋白質表面への露出度や存在環境を示す。アミノ酸配列について用いられる「ギャップ」という用語は、2以上のアミノ酸配列間の対応付けにおいていずれか一方の配列に対応するアミノ酸残基がない部分を指し、一方の配列から見ると1又は2個以上のアミノ酸残基の挿入及び／又は欠失していることを意味する。

本発明の好ましい態様では、このデータベースに参照蛋白質毎に立体構造を反映して分割した2以上の部分配列の情報とアミノ酸残基毎の環境情報とを格納しておく。参照蛋白質の環境情報と質問配列中の対応アミノ酸残基の疎水度パラメータとから計算されるマッチングスコアに基づいて2つの配列をマッチングする。疎水度パラメータは20種のアミノ酸残基に対して予め数値化しておくことが望ましい。

部分配列のうち、疎水コアの形成に関与するコア部分配列についてはギャップを入れずに質問配列上をスライドさせ、両端においてのみアミノ酸残基の増減(隣接するサブ部分配列の末端から1又は2個以上のアミノ酸残基をとて該コア部分配列に組み込むか、又はその逆の操作を意味する)を考慮してマッチングを行い、スコアのよい鋳型候補蛋白質を選別する。次に、疎水コアの安定化に関与しないサブ部分配列について必要に応じてギャップを考慮したマッチングを行い、鋳型候補の数を絞る。最終的な鋳型蛋白質の選択は、各蛋白質の最適マッチングのスコアを自己マッチングスコアで規格化して比べることにより行うことができる。

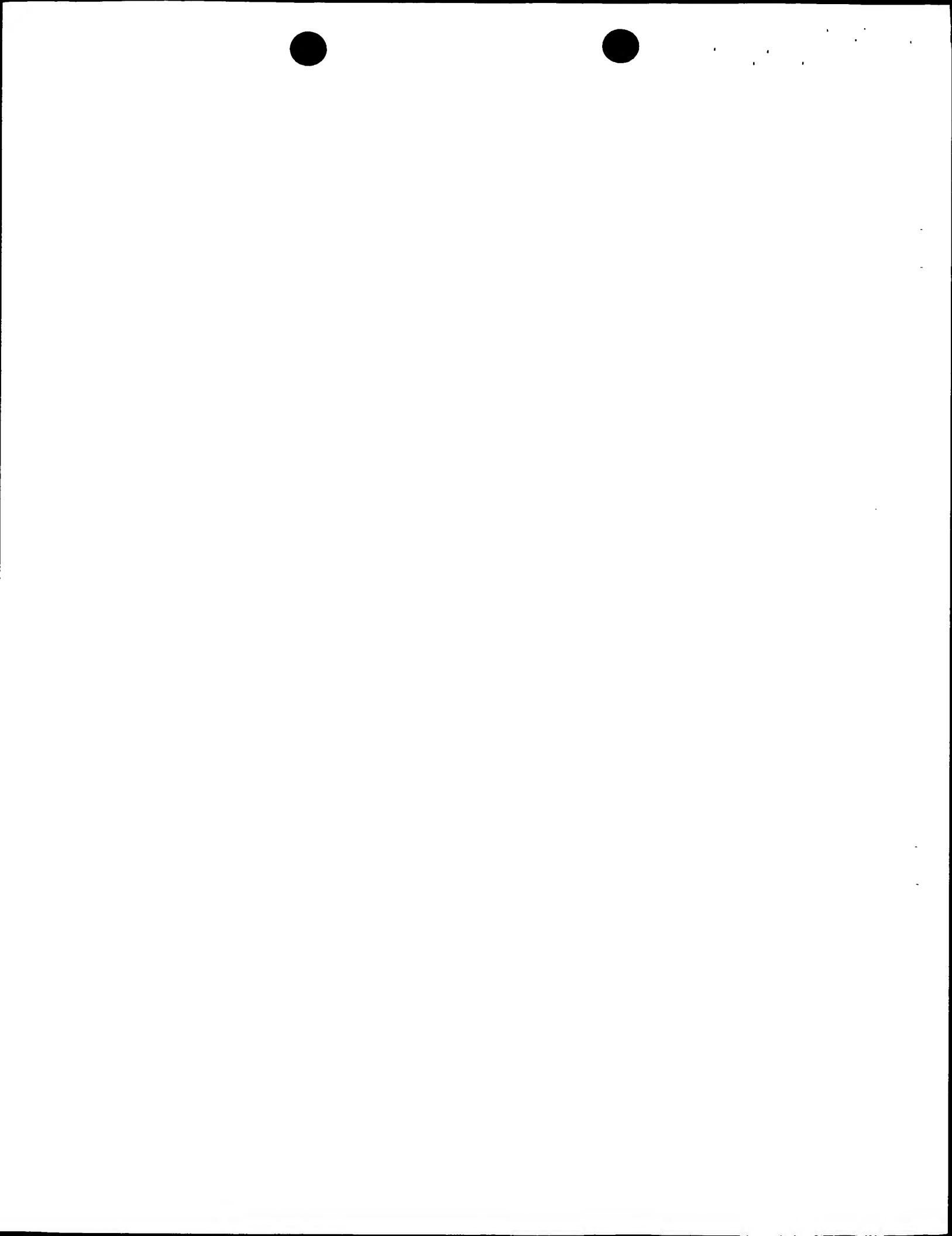
本発明の方法は、マッチングスコアが高いほど2つの配列間でスキヤッフォールドや立体構造の類似性が高いとの仮定に基づいており、配列の相同性が殆どない配列間の残基のマッチングを蛋白質の折り畳み原理に従って行うことによって、参照蛋白質からの鋳型蛋白質の適切な選択を可能にしている点に特徴がある。ある。本発明の好ましい態様では、(1) データベースの作成；(2) 部分配列を用いたマッチング；(3) マッチングスコアの計算；(4) 参照蛋白質から鋳型候補蛋白質の選択；(5) 鋳型蛋白質の選択を主要な要素としている。



一般的な水溶性の単一サブユニットの蛋白質を構成するペプチド鎖については、アミノ酸残基の疎水性側鎖ができるだけ分子内部に埋もれて露出せず、一方、親水性側鎖はできるだけ蛋白質分子表面に露出するのが自由エネルギー的に有利である。もっとも、細胞膜に相互作用する蛋白質や複数のサブユニットからなる蛋白質など存在環境が異なる蛋白質では、個々の蛋白質構造やサブユニット構造だけをみると、疎水性側鎖が蛋白質表面に露出していることもある。本発明の方法では、このような蛋白質の立体構造の多様性も考慮されており、個々のアミノ酸残基の存在環境を反映したスコア付けができるようになっている。アミノ酸残基毎の特定2次構造のとり易さの概念（例えば P. Y. Chou, & G. D. Fasman, *Adv. enzymol.* 47, 45, 1978）は基本的には用いないが、必要に応じてそれらの概念を加味したスコアを用いることができる。

### 実施例

以下、本発明の方法の好ましい態様をスキームで示し、このスキームに従って本発明の方法をより具体的に説明するが、本発明の方法はこのスキームの細部又は以下の説明の細部に限定されることはない。



立体構造が既知または推定可能な蛋白質  
(参照蛋白質) のアミノ酸配列と三次元座標

コア部分配列とサブ部分配列への分割  
アミノ酸毎に環境情報アサイン  
(露出度パラメータと存在環境)  
自己マッチングスコア計算  
その他の情報アサイン

部分配列情報、アミノ酸毎のパラメータ ←  
含む参照蛋白質のデータベース

質問配列 ←→ 参照蛋白質を1個とりだす

コア部分配列でのマッチング  
各配列の両端での残基の増減

全コア部分配列の最適マッチング

全参照蛋白質間でスコアの比較

鋳型候補蛋白質 (数個～数十個)  
(推定スキヤッフォールド)

サブ部分配列のマッチング

鋳型蛋白質 (1～数個)  
(推定スキヤッフォールド)

側鎖の置換、ギャップの処理

立体構造の構築 (モデリング)

構造最適化  
エネルギーの比較、機能面の考察

立体構造の推定



## (A) データベースの作成

立体構造が既知または推定可能な参照蛋白質について、立体構造に関する情報として、各アミノ酸残基の環境情報及び部分配列の情報を含むデータベースを作成しておく。データベースに収納する蛋白質としては、立体構造の情報が既知または推定可能な蛋白質であれば、すべてエントリーとすることができます。同時に構造決定された蛋白質中に複数のサブユニットが含まれるときは、独立のエントリーとしておくことが好ましい。ペプチド鎖で繋がった複数ドメインからなる構造については、全体構造とともに各ドメインも独立エントリーとしておくことが好ましい。

参照蛋白質毎に含める情報は以下の通りである。

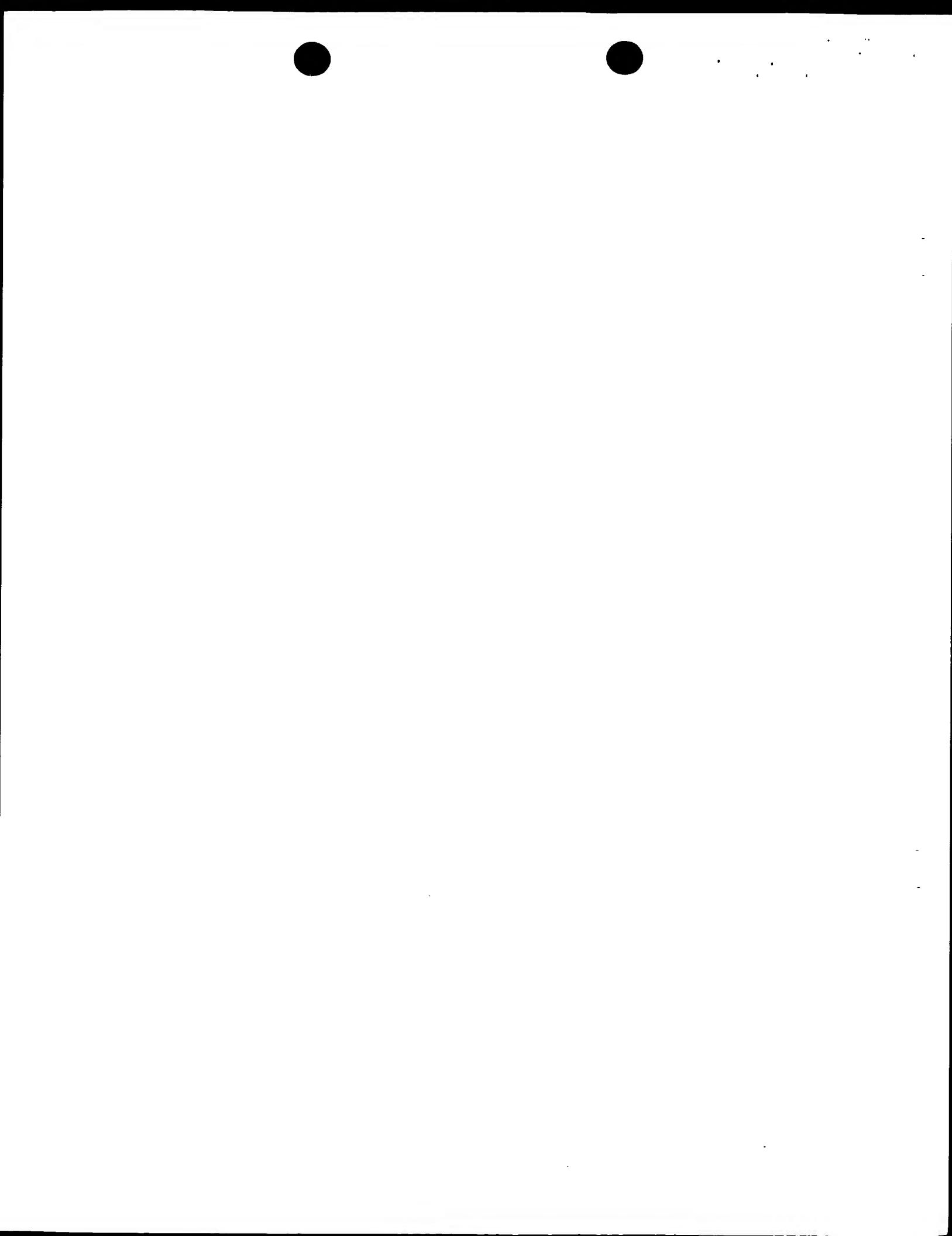
### (1) 一般的な情報

- (a) 蛋白質名（蛋白質コード）、サブタイプ、アミノ酸数、アミノ酸配列、ドメイン、サブユニットなどに関する情報。
- (b) 立体構造に関する情報として、立体構造の決定（又は推定）方法（結晶解析法、n m r 法、又はモデリング法のいずれにより立体構造を決定したか）、P D B コード、モデリング法による場合には鋳型として利用した蛋白質分子名、結晶解析の場合には共結晶化された分子がある場合にはその化学名などの情報のうち適宜のものを含めておく。
- (c) 生物学的意義に関する情報として、生物学的機能、生物種、存在する組織・臓器、エフェクター分子などの情報を含めておく。

### (2) 部分配列情報

蛋白質の立体構造における特徴に従って、配列を分割して2以上の部分配列とし、各部分配列について以下の情報をデータベースに含めるのが望ましい。

- (a) N末端から何番目の部分配列か
- (b) コア部分配列かサブ部分配列かのフラグ
- (c) 始点と終点（N末端からのアミノ酸配列番号で）、配列の長さ、その距離及



びベクトル、サブ部分配列ならば最短の残基数

(d) 部分配列間の距離及びベクトル

(e)  $\beta$  -シート形成の水素結合をする相手の部分配列番号、および逆平行または平行  $\beta$  -シートなどの区別、あるいは一定の距離内にある部分配列番号など

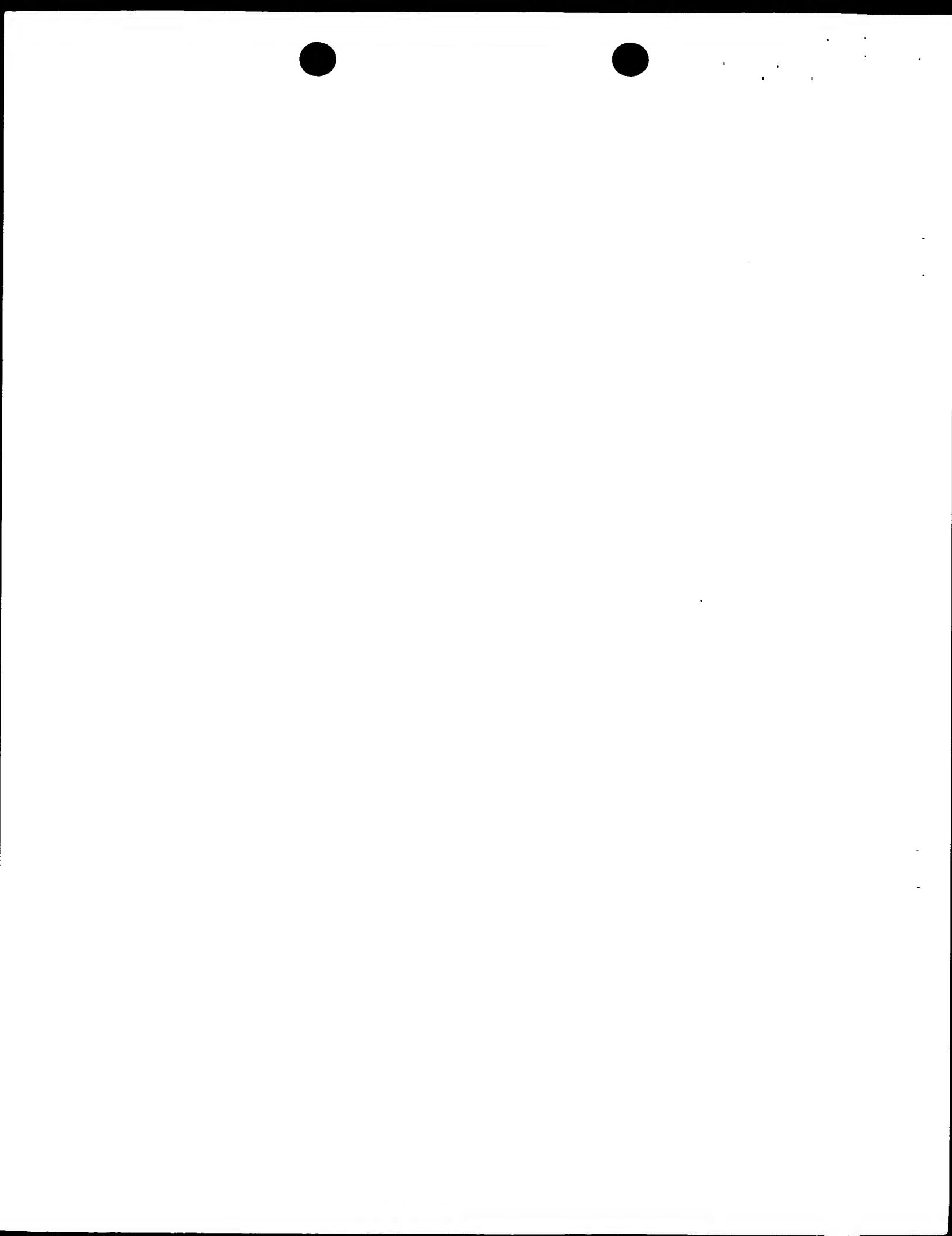
(f) 自己マッチングスコア (後述)

部分配列への分割の基準は特に限定されないが、 $\alpha$  -ヘリックス構造や  $\beta$  -ストランド構造などの2次構造を形成し実質的に疎水コアの形成に関わる部分をコア部分配列とし、それ以外の部分をサブ部分配列とすることができます。各部分配列は、例えば、7残基以上を含む連続した配列とすることが望ましい。 $\beta$  -ターン構造については、はじめからコアまたはサブ部分配列に加えても構わないし、別のフラグによって区別し、検索時にコアかサブかの判断を加えてもよい。分割の作業は、コンピュータグラフィックス画面上で対話的に個々の蛋白質について行ってもよいし、分割の基準を定めたプログラムを作成して自動的に行ってもよい。その判断基準として、例えば、隣り合う4つのCアルファ原子のなす擬ねじれ角などの数値を基準に用いれば、自動的に部分配列に分割できる。

(3) アミノ酸残基毎の環境情報

(a) 露出度パラメータ

参照蛋白質の全てについて、立体構造に基づいて残基毎の蛋白質表面への側鎖の露出度及び蛋白質内部への側鎖の埋没度を計算し、その計算値に基づいて露出度パラメータを割り振る。本明細書において用いられる「露出度パラメータ」という用語は、各残基の側鎖がどの程度分子表面に露出しているか、又は埋没しているかを示す数値を意味している。露出度パラメータを定義する方法は特に限定されず、いかなる方法を採用してもよいが、露出度が高いものはマイナスの値、露出度が低いものはプラスの値をとるように設定するのが望ましい。例えば、立体構造において各アミノ酸側鎖の溶媒への接触可能表面と蛋白質原子への接触表面積を計算して、その差から露出度パラメータを算出して利用することができる。



また、例えば、全分子表面に対する溶媒接触表面の割合を基準として定めることもできる。

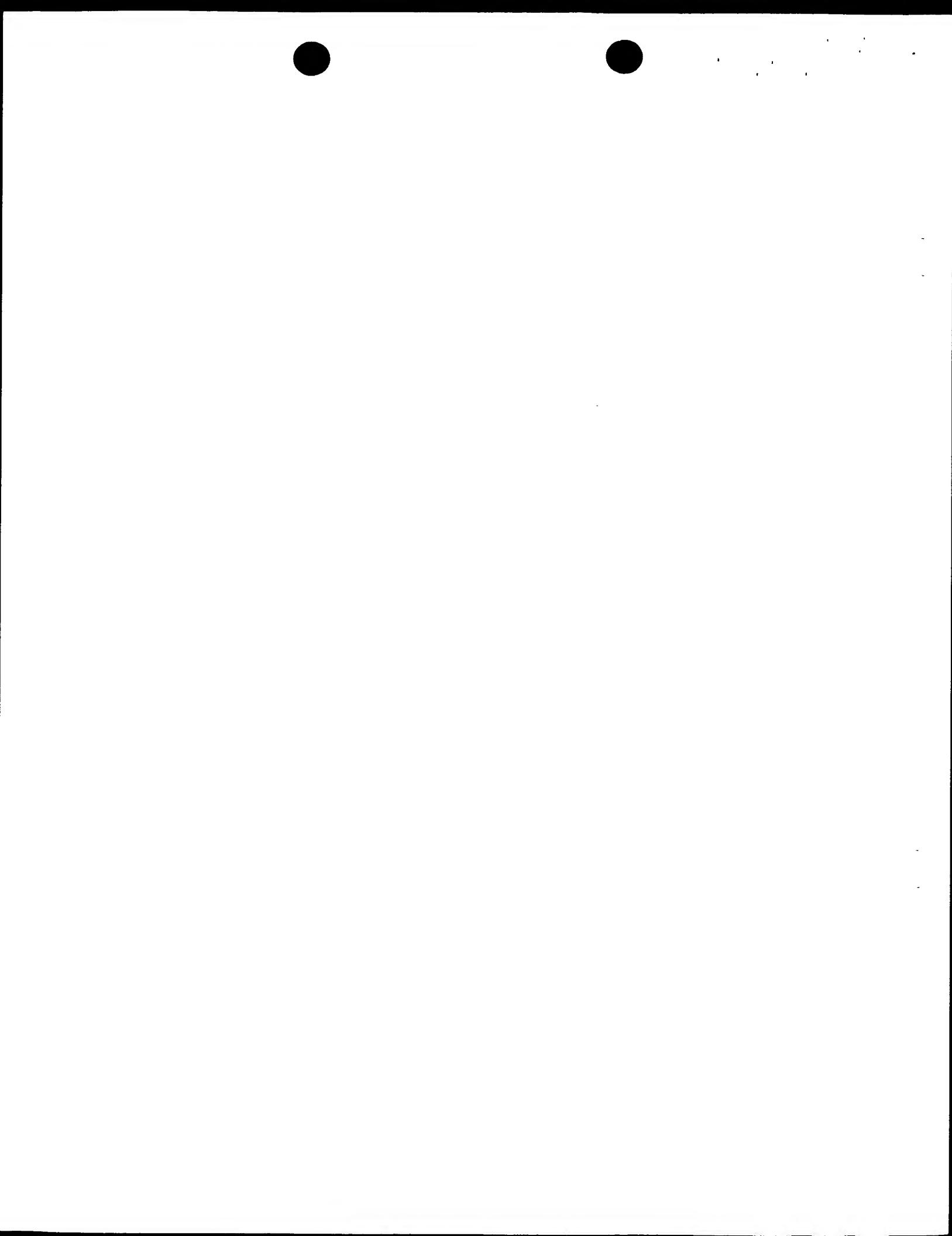
#### (b) 存在環境フラグの設定

細胞膜に相互作用する蛋白質では膜と相互作用する部分の表面に疎水性アミノ酸残基の側鎖が露出しており、また、サブユニットやドメイン構造が集合して安定化する接触面を有する蛋白質では、接触面に疎水性アミノ酸残基の側鎖が露出している場合があり、これらの蛋白質は単独で存在する水溶性蛋白質とは異なる性質を有している。このような蛋白質については、一般的には、先述の折り畳み原理はそのままでは適用できない。データベース作成時に、例えば、次のような存在環境フラグを露出度パラメータとは別に各アミノ酸残基に与えておくことができる。

例えば、蛋白質の由来や機能に関する実験結果および立体構造から、各アミノ酸の存在環境が以下のいずれに相当するかを推定して環境フラグを与えておき、マッチングやスコアの計算の際に考慮することが可能である。

- 0 : 不明 (未定義または定義できず)
- 1 : 分子内部 (蛋白質内、サブユニット内、ドメイン内の接触)
- 2 : 分子内孔 (リガンド結合部位)
- 3 : 分子表面 (水環境と接触)
- 4 : 分子表面 (別蛋白質、別サブユニット、別ドメインと接触)
- 5 : 分子表面 (膜と接触)

さらに、立体構造形成に特殊な影響を与えるアミノ酸残基については、特殊残基であることを示すフラグを与えておき、マッチングやスコアの計算の際に考慮することができる。例えば、S-S結合しているシステイン残基やプロリンのように主鎖に水素結合性官能基が不足しているアミノ酸残基、または側鎖原子間に強い親水性相互作用を形成する可能性のある残基などに適用することができる。



## (B) 質問配列と参照蛋白質の配列間のマッチング

一方の配列を他方の配列上スライドさせて最適のマッチングを効率よく探すために、部分配列の概念を利用することができる。そのために、上記データベースは、各参照蛋白質について立体構造から得られる部分配列の情報とアミノ酸残基ごとの環境情報がアミノ酸配列順に収めてある。一方、質問配列が有している情報はアミノ酸配列の情報だけであり、アミノ酸残基毎の疎水度パラメータの表からとった値を当てはめてスコアの計算に用いる。データベースから参照蛋白質を1つずつ取り出し、アミノ酸配列中に出現する順に並べた部分配列群を質問配列上でスライドさせて、部分配列群と質問配列との間でマッチングスコアの最もよいマッチングを探す。

### (1) 部分配列を用いたマッチング

部分配列を用い、かつコア部分配列ではギャップを考慮しないでアミノ酸残基単位でマッチングを行うことにより、最適マッチングの配列間対応付けを高速に検索することができる。一般に、進化の過程では残基の置換とともに挿入や欠失が起きることが多く、その考慮は配列のマッチングに不可欠である（「従来の技術」の欄を参照）。しかしながら、一般的には、ギャップが入るのはサブ部分配列であることが多い。これは、疎水コアの安定化に関するコア部分配列でその両端以外で挿入や欠失が起きると、その安定スキヤッフオールド自体が損なわれ、蛋白質の立体構造が大きく変化してしまうからである。

そこで、本発明の方法の好ましい態様では、コア部分配列とサブ部分配列とに分けて2段階のマッチングを行う。第一段階のコア部分配列を用いたマッチングにおいては、まずギャップを考慮せずに1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の両端においてのみアミノ酸残基の増減を考慮したマッチングを行って最適マッチングを探索する。

各コア部分配列について質問配列の上をスライドさせながらマッチングスコアを計算して保存する。全部のコア部分配列について同様に計算した後、全体とし



て最適なマッチングを決定する。2以上のコア部分配列を用いる場合には、質問配列上に2以上のコア部分配列を参照蛋白質のアミノ酸配列中での出現順序に従って重なり合わないように配置し、コア部分配列間には4個程度以上のアミノ酸残基の存在を仮定して（立体的に結合可能なアミノ酸残基数で隣のコア部分配列と順番に繋がるという条件、例えば $\beta$ -ターンなどに要するアミノ酸残基数）、それぞれのコア部分配列の順序を変えずに質問配列上をスライドさせ、最も総スコアのよいマッチングを選択する。この際、各コア部分配列のマッチングスコアが最大である必要はない。この第一段階で、高いスコアを与える鑄型蛋白質を数個から数十個選択し、スキヤッフォールド候補として第二段階に進む。

#### (C) マッチングスコア

マッチングスコアは参照蛋白質の環境情報の露出度パラメータ  $EP(i)$  と質問配列の残基の疎水度パラメータ  $HB(j)$  を用いて計算する。 $i$  は参照蛋白質のアミノ酸配列中のアミノ酸残基番号であり、 $j$  はそれと対応づけられた質問配列の残基番号を指す。マッチングスコアの計算式は、参照蛋白質の分子内部に埋没した側鎖環境に質問配列の疎水性の強いアミノ酸残基が対応し、分子表面に露出した側鎖環境に質問配列の親水性の強いアミノ酸残基が対応すると高いスコアが得られるような計算式であれば、いかなるものを利用してもよい。マッチングスコアは例えば、最も単純には次式によって計算することができる。

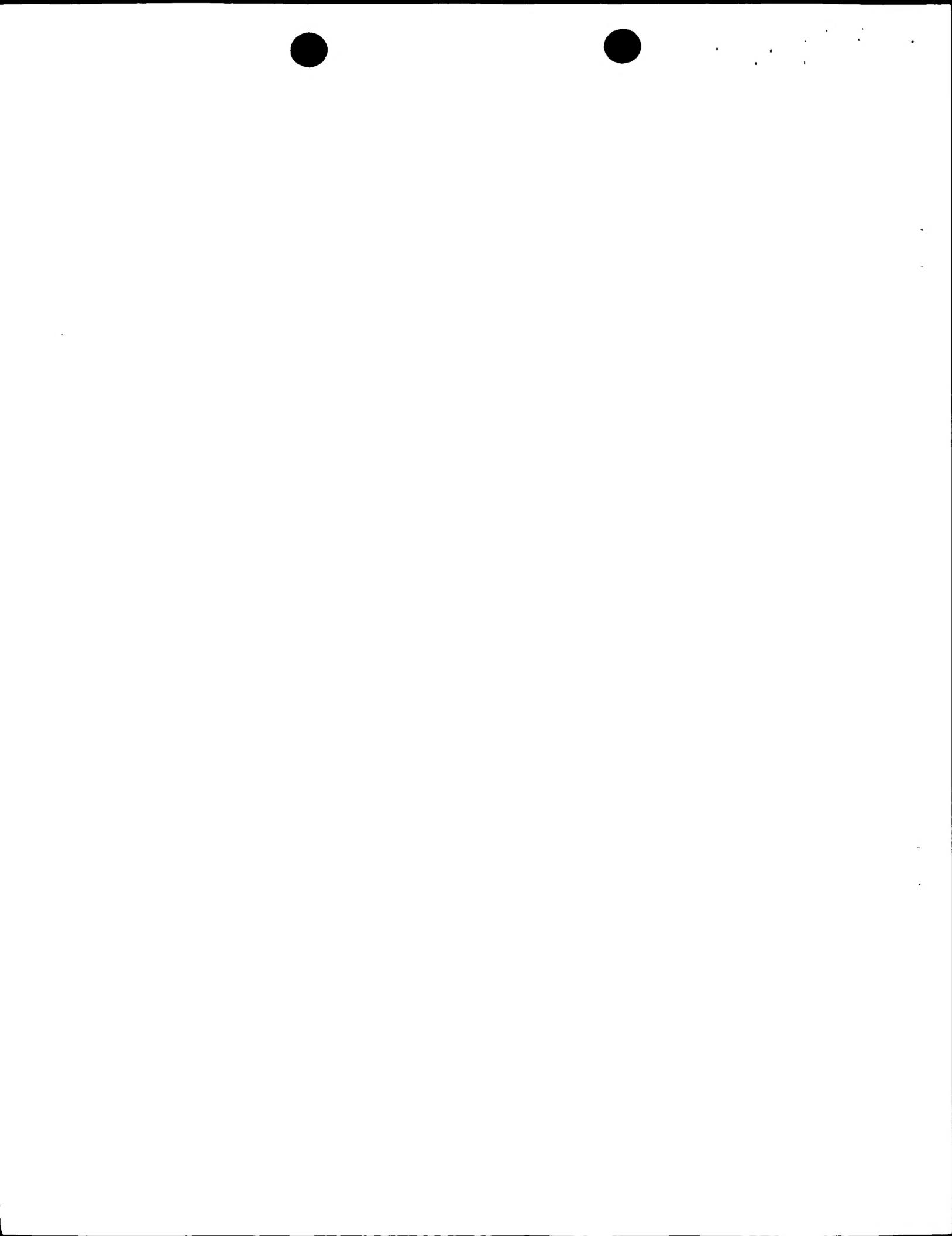
残基毎のマッチングスコア =  $EP(i) \times HB(j)$

部分配列のマッチングスコア = 配列に含まれる残基のマッチングスコアの和

全配列のマッチングスコア = 全部分配列のマッチングスコアの和

##### (1) 疎水度パラメータ

20種のアミノ酸残基のそれぞれに疎水性又は親水性の性質に関連した疎水度パラメータを与えておく。疎水度パラメータの決定方法は特に限定されず、いかな



る基準による値を用いてもよい。例えば、文献に記載されたアミノ酸毎の疎水性値を用いてもよく、又は適宜の方法により独自の基準で算出したものを用いてもよい。また、あるアミノ酸について、結晶解析された蛋白質中の全出現回数に対して蛋白質分子内部に埋没された残基の比率を統計的に求めておき、その比率を該アミノ酸の疎水度パラメータとして利用してもよい。

また、例えば、個々のアミノ酸残基に別々の値を与えてよいが、次のように段階化した疎水度パラメータを与えることもできる。

表 1

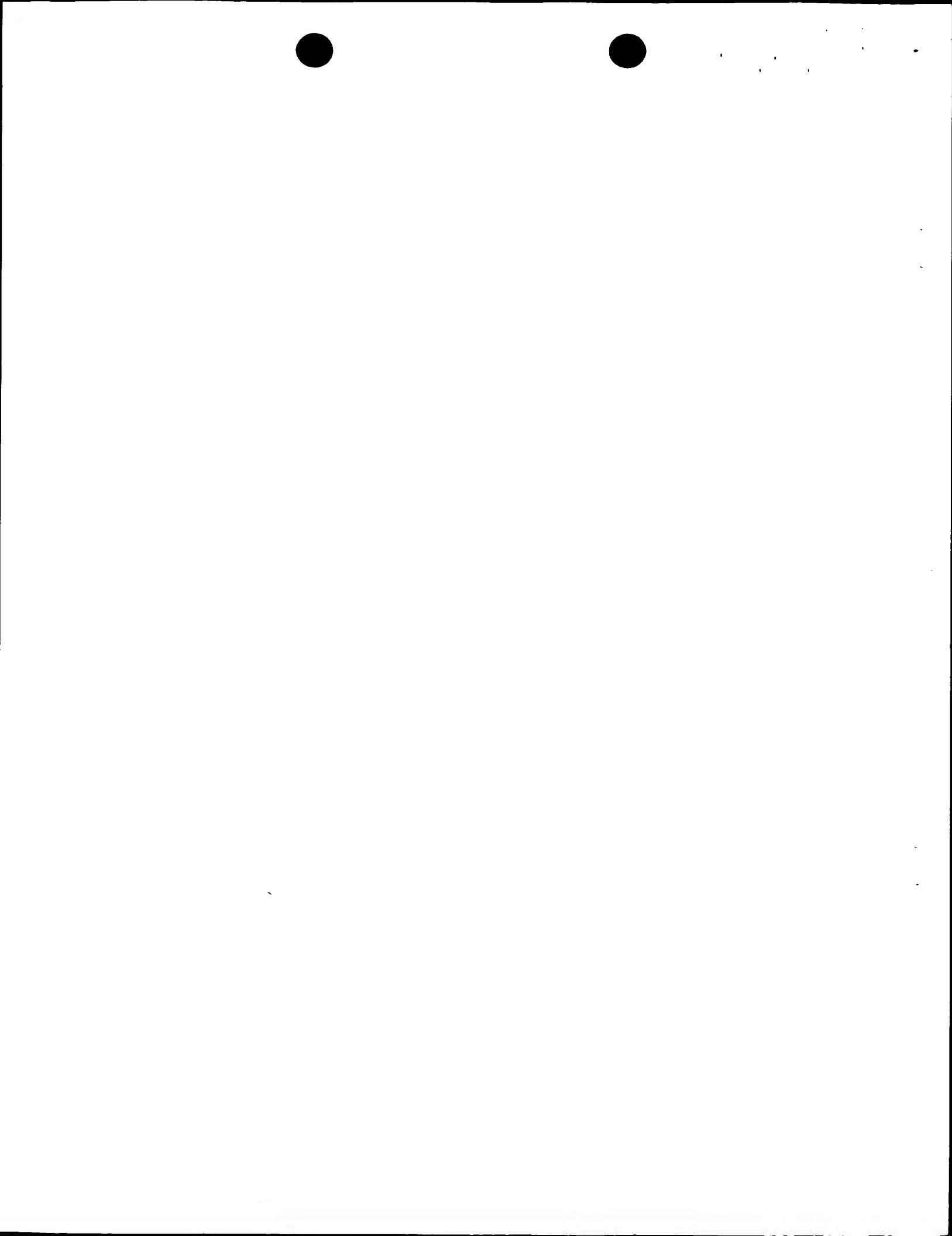
- 2 : 強い疎水性 (イソロイシン、バリン、ロイシン、フェニルアラニン)
- 1 : 弱い疎水性 (アラニン、メチオニン、シスチン、チロシン)
- 0 : ほぼ中性 (グリシン、プロリン、リジン、アルギニン)
- 1 : 弱い親水性 (スレオニン、ヒスチジン)
- 2 : 強い親水性 (セリン、アスパラギン、アスパラギン酸、グルタミン、グルタミン酸)

## (2) 自己マッチングスコア

アミノ酸数及びアミノ酸組成の異なる蛋白質間で質問配列へのマッチングの良さを比較するためにはスコアの規格化をしておくことが望ましい。そのために、各参照蛋白質について、それ自体のアミノ酸配列の露出度パラメータ EP(i) とデータベースに用意された環境情報の疎水度パラメータ HB(i) とからマッチングスコアを予め計算してデータベースに保存しておく。例えば、次式のように計算するとよい。

$$\text{自己マッチングスコア} = \sum (EP(i) \times HB(i))$$

全部分配列の質問配列へのマッチングが終了して最適マッチングが得られた後、



得られたマッチングスコアに自己マッチングスコアを乗じて規格化することができる。全参照蛋白質について規格化された最適マッチングスコア同士を比較して最適な鋳型候補蛋白質を選択することができる。自己マッチングスコア及びマッチングスコアは、アミノ酸残基数が多いほど大きな値を取りやすい。

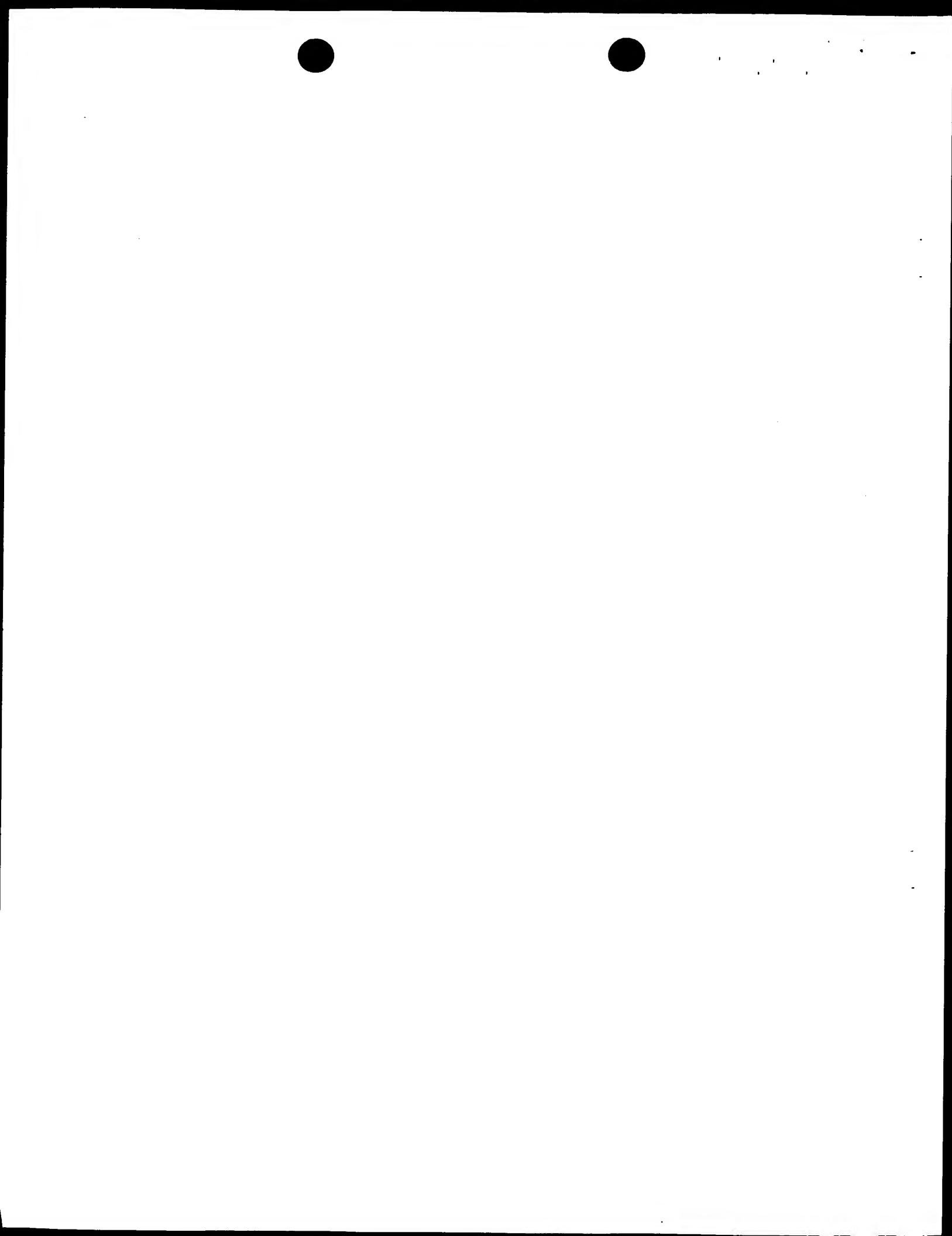
#### (D) 鋳型候補蛋白質の選択

鋳型候補蛋白質の選択の手順は概ね以下の通りである。

- (a) データベースから参照蛋白質を1個づつ取り出し、質問配列に対しマッチングを行う；
- (b) コア部分配列につきギャップを考慮せずに質問配列上をスライドしマッチングスコアを算出する；
- (c) 必要に応じて、各コア部分配列のN末端またはC末端の残基を増減しながらマッチングし、最適マッチングを得る；
- (d) 参照蛋白質全部について工程(a)～(c)を行い、最適マッチングとマッチングスコアを得る；
- (e) 規格化したスコアにより参照蛋白質からスコアのよいものを鋳型候補蛋白質とする（この段階で、それらの構造は質問配列のスキヤッフォールドの候補とすることができる）；
- (f) コア部分配列間をつなぐサブ部分配列のマッチングを行う。質問配列の対応する配列部分との配列長の違いやギャップの存在を考慮し、最適のマッチングとマッチングスコアを得る；及び
- (g) 規格化したスコアによって鋳型蛋白質を選択する。

#### (E) 立体構造の構築

質問配列の蛋白質の立体構造は、鋳型蛋白質のスキヤッフォールドの構造、及び該参照蛋白質と質問配列との最適マッチングの結果に基づいて、アミノ酸残基の側鎖の置換を行うことにより構築することができ、質問配列に対応した3次元座

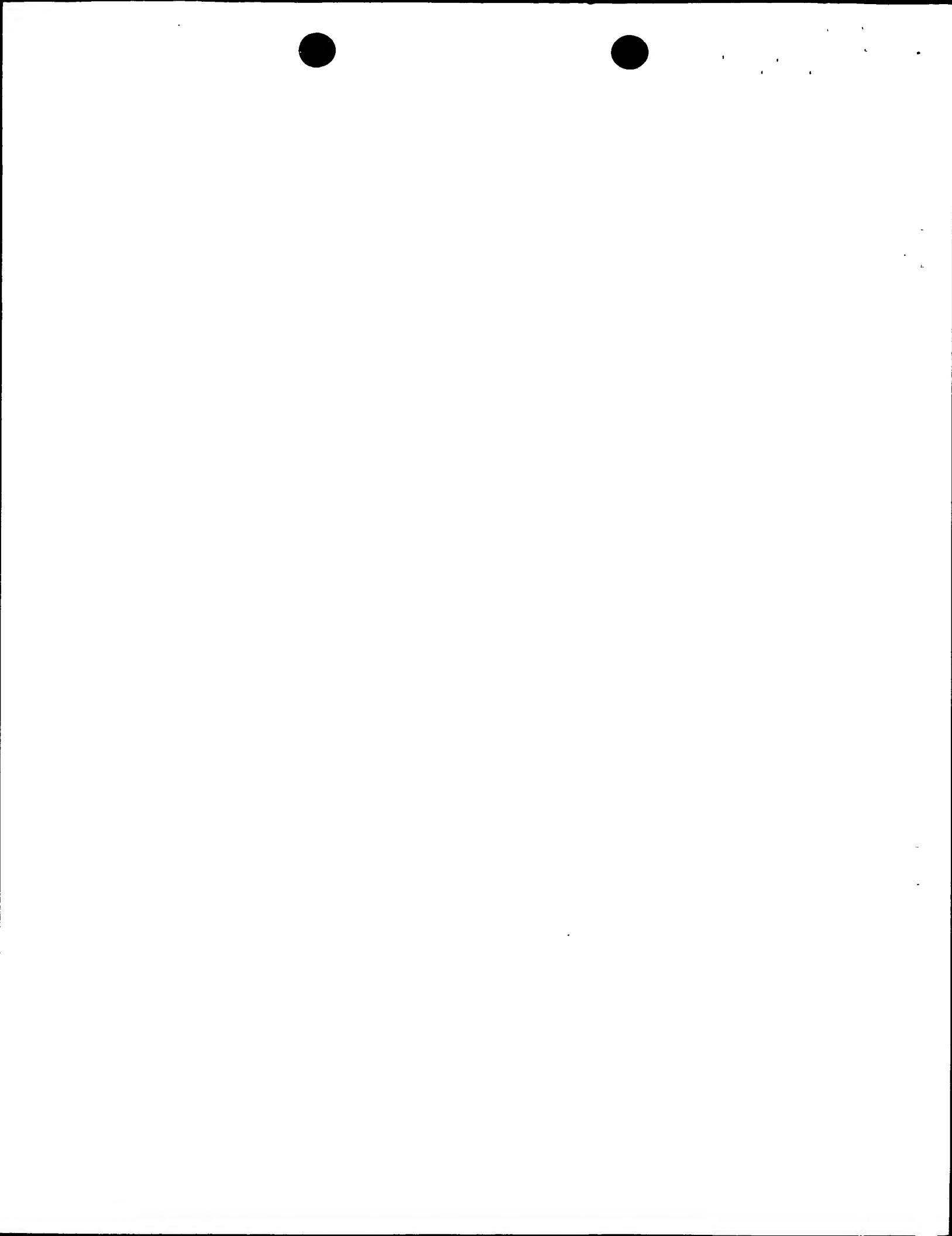


標を得ることができる。優劣をつけ難い2以上の鋳型蛋白質がある場合には、すべてについて立体構造を構築することが望ましい。サブ部分配列の長さが鋳型候補と異なる場合には、結晶構造に現れたループ構造を集めたデータベース等を用いて、該サブ部分配列の適切なトポロジーを決定することができる。鋳型のスキヤッフォールドが修正される部分については、マッチングスコアと同様に残基の性質と露出度を考慮しつつトポロジーを決定することができる。重大な立体障害や立体構造を不安定化する分子内部の隙間などの有無を確認し、また構造最適化計算や分子動力学計算により構造の微調整を行ったのちに立体構造の安定性を比較する。

最後に、全ての鋳型に基づいて構築された立体構造をエネルギーとマッチングスコアにより順位付けするが、該質問配列の蛋白質の機能が既知の場合には、該鋳型蛋白質に知られた機能との対応、その機能の発現に関与するものと推定されるアミノ酸残基の立体構造上の位置の妥当性、さらにはアミノ酸変異による機能への影響などの情報を鋳型蛋白質の選別に利用できる。

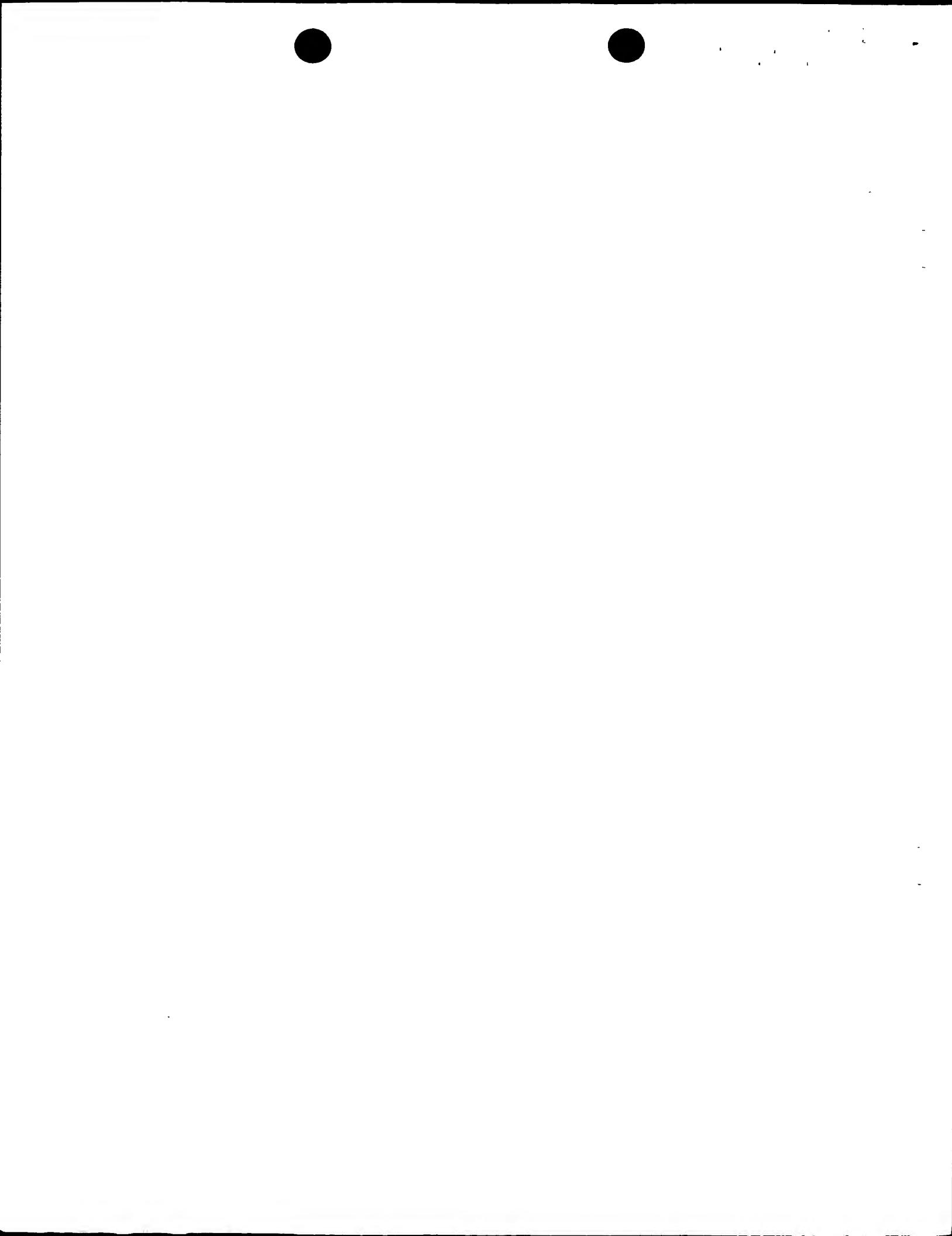
#### 産業上の利用可能性

本発明の方法によれば、立体構造が既知または推定可能な蛋白質のアミノ酸配列データベースに基づいて、アミノ酸配列の情報のみからそのアミノ酸配列により構築される蛋白質の立体構造に関する情報を信頼性高く効率的に入手することができる。



## 請求の範囲

1. 立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキップオールドを推定する方法。
2. 参照蛋白質のアミノ酸配列が該参照蛋白質の立体構造に基づいて連続した2以上のアミノ酸残基からなる2以上の部分配列に分割された請求の範囲第1項に記載の方法。
3. 参照蛋白質のアミノ酸配列が疎水コアの形成に実質的に関与する1又は2以上のコア部分配列と疎水コアの形成に実質的に関与しない1又は2以上のサブ部分配列とに分割された請求の範囲第2項に記載の方法。
4. 参照蛋白質における各アミノ酸残基の側鎖の蛋白質内部への埋没度及び／又は蛋白質表面への露出度の情報と、質問配列の各アミノ酸の疎水性及び／又は親水性の性質とに基づいてマッチングを行う請求の範囲第1項ないし第3項のいずれか1項に記載の方法。
5. 参照蛋白質の1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の片端又は両端以外においてはギャップを考慮せずにマッチングを行う請求の範囲第1項ないし第4項のいずれか1項に記載の方法。
6. ギャップが1又は2以上のアミノ酸残基の削除又は付加である請求の範囲第5項に記載の方法。
7. マッチングが以下の工程：
  - (a) 1又は2以上のコア部分配列を質問配列上でスライドさせながら、必要に応じて該コア部分配列の片端又は両端においてはギャップを考慮してマッチングを行う工程（ただし、2以上のコア部分配列を用いる場合には、該コア部分配列は



参照蛋白質のアミノ酸配列での出現順に順番に配置する) ; 及び

(b) 工程(a) に続いて、必要に応じてギャップを考慮しつつサブ部分配列を質問配列上でスライドさせる工程を含む請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

8. 参照蛋白質のアミノ酸残基の側鎖に関する環境情報と質問配列上の対応アミノ酸残基の疎水度パラメータとから計算したスコアに基づいて最適マッチングを選択する、請求の範囲第1項ないし第7項のいずれか1項に記載の方法。

9. 参照蛋白質についての自己マッチングスコアを用いて上記スコアを規格化する工程を含む請求の範囲第8項に記載の方法。

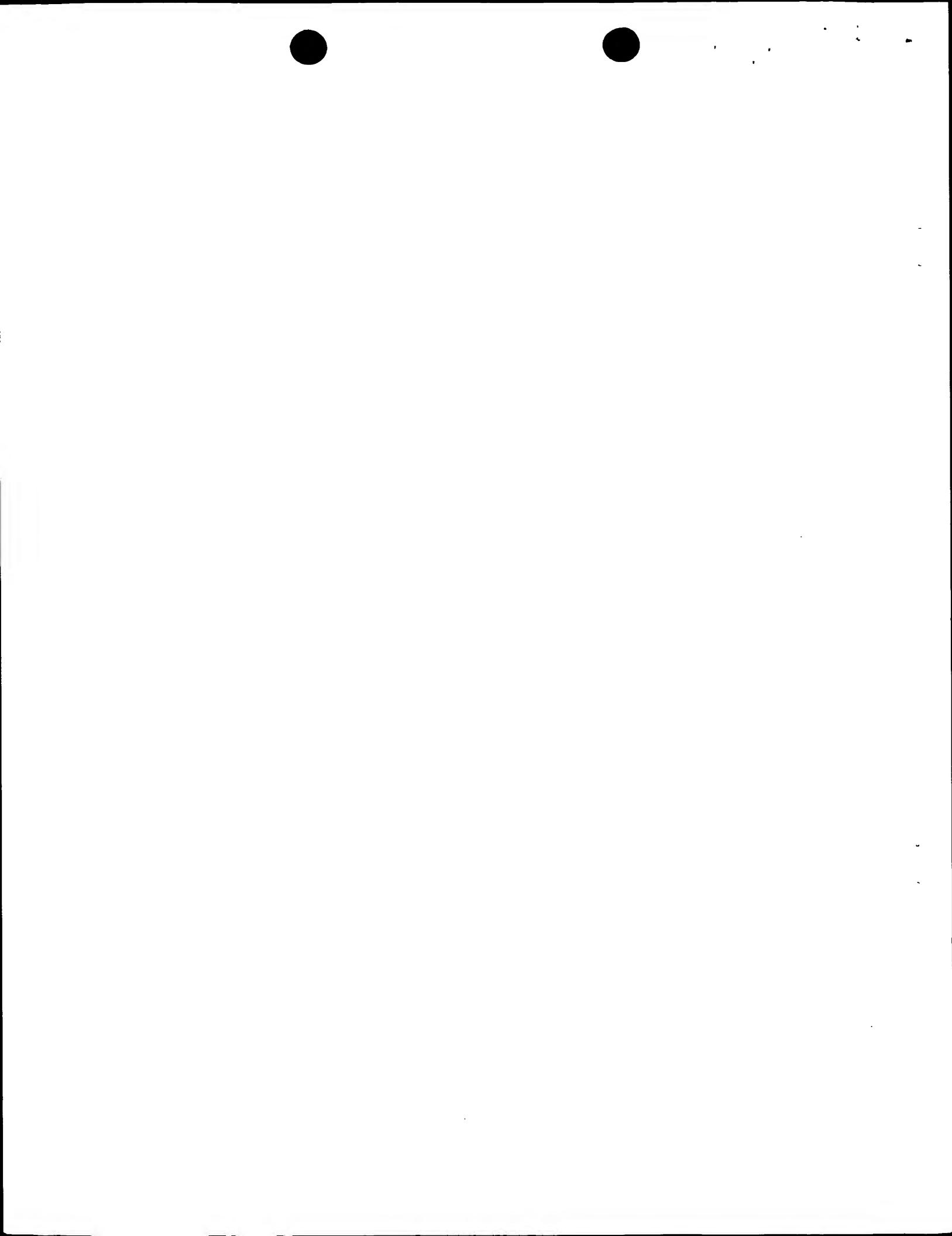
10. さらに質問配列の蛋白質の立体構造を構築する工程を含む請求の範囲第1項ないし第9項のいずれか1項に記載の方法。

11. 立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースであって、請求の範囲第1項ないし第10項のいずれか1項の方法に用いるためのデータベース。

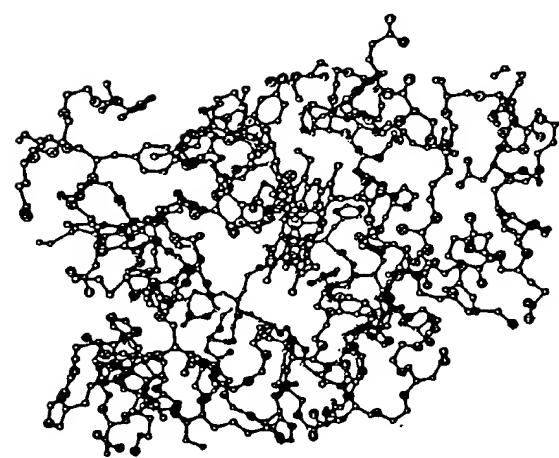
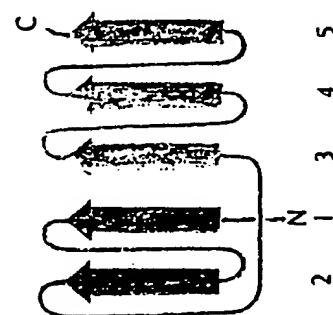
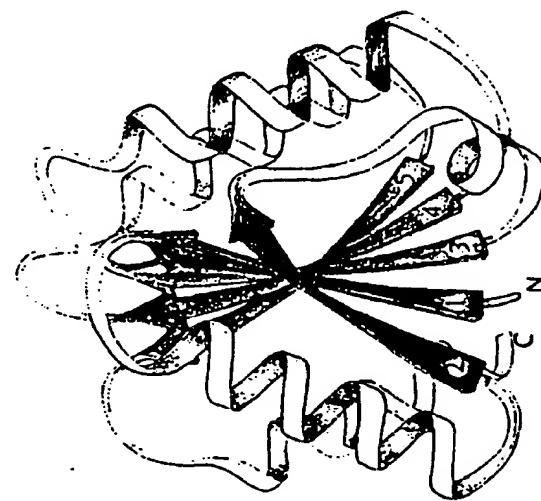


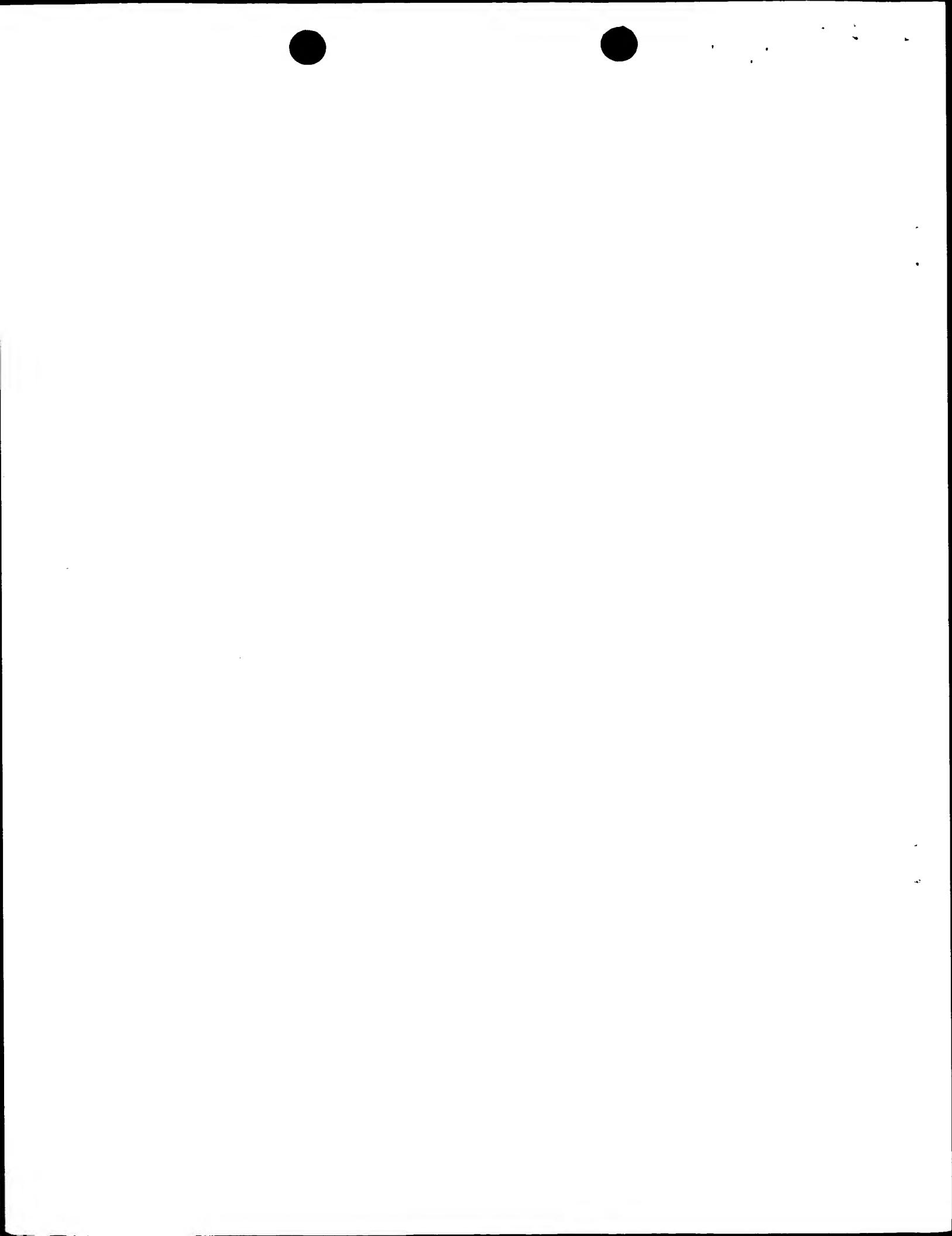
## 要 約 書

立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列を疎水コアの形成に実質的に関与するコア部分配列と関与しないサブ部分配列に分割し、各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、部分配列毎に参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鑄型蛋白質を選択して質問配列の蛋白質のスキヤッフォールドを推定する方法。



第1図





## 特許協力条約

PCT

## 国際予備審査報告

(法第12条、法施行規則第56条)  
[PCT36条及びPCT規則70]

出願人又は代理人 の書類記号 98215M	今後の手続きについては、国際予備審査報告の送付通知（様式PCT/IPEA/416）を参照すること。	
国際出願番号 PCT/JP98/04457	国際出願日 (日.月.年) 02.10.98	優先日 (日.月.年) 02.10.97
国際特許分類 (IPC) Int. Cl. G01N33/68, G06F17/00, G06F15/18, C07K14/00		
出願人（氏名又は名称） 板井昭子		

1. 国際予備審査機関が作成したこの国際予備審査報告を法施行規則第57条（PCT36条）の規定に従い送付する。
2. この国際予備審査報告は、この表紙を含めて全部で <u>4</u> ページからなる。
<input type="checkbox"/> この国際予備審査報告には、附属書類、つまり補正されて、この報告の基礎とされた及び／又はこの国際予備審査機関に対して訂正を含む明細書、請求の範囲及び／又は図面も添付されている。 (PCT規則70.16及びPCT実施細則第607号参照) この附属書類は、全部で <u>      </u> ページである。
3. この国際予備審査報告は、次の内容を含む。
I <input checked="" type="checkbox"/> 国際予備審査報告の基礎 II <input type="checkbox"/> 優先権 III <input checked="" type="checkbox"/> 新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成 IV <input type="checkbox"/> 発明の単一性の欠如 V <input checked="" type="checkbox"/> PCT35条(2)に規定する新規性、進歩性又は産業上の利用可能性についての見解、それを裏付けるための文献及び説明 VI <input type="checkbox"/> ある種の引用文献 VII <input type="checkbox"/> 国際出願の不備 VIII <input type="checkbox"/> 国際出願に対する意見

国際予備審査の請求書を受理した日 02.10.98	国際予備審査報告を作成した日 30.04.99
名称及びあて先 日本国特許庁 (IPEA/JP) 郵便番号 100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官（権限のある職員） 亀田宏之 電話番号 03-3581-1101 内線 3252
	2J 9015



## I. 国際予備審査報告の基礎

1. この国際予備審査報告は下記の出願書類に基づいて作成された。（法第6条（PCT14条）の規定に基づく命令に応答するために提出された差し替え用紙は、この報告書において「出願時」とし、本報告書には添付しない。PCT規則70.16, 70.17）

 出願時の国際出願書類

<input type="checkbox"/> 明細書 第 _____	ページ、	出願時に提出されたもの
<input type="checkbox"/> 明細書 第 _____	ページ、	国際予備審査の請求書と共に提出されたもの
<input type="checkbox"/> 明細書 第 _____	ページ、	付の書簡と共に提出されたもの
<input type="checkbox"/> 請求の範囲 第 _____	項、	出願時に提出されたもの
<input type="checkbox"/> 請求の範囲 第 _____	項、	PCT19条の規定に基づき補正されたもの
<input type="checkbox"/> 請求の範囲 第 _____	項、	国際予備審査の請求書と共に提出されたもの
<input type="checkbox"/> 請求の範囲 第 _____	項、	付の書簡と共に提出されたもの
<input type="checkbox"/> 図面 第 _____	ページ/図、	出願時に提出されたもの
<input type="checkbox"/> 図面 第 _____	ページ/図、	国際予備審査の請求書と共に提出されたもの
<input type="checkbox"/> 図面 第 _____	ページ/図、	付の書簡と共に提出されたもの
<input type="checkbox"/> 明細書の配列表の部分 第 _____	ページ、	出願時に提出されたもの
<input type="checkbox"/> 明細書の配列表の部分 第 _____	ページ、	国際予備審査の請求書と共に提出されたもの
<input type="checkbox"/> 明細書の配列表の部分 第 _____	ページ、	付の書簡と共に提出されたもの

2. 上記の出願書類の言語は、下記に示す場合を除くほか、この国際出願の言語である。

上記の書類は、下記の言語である \_\_\_\_\_ 語である。

- 国際調査のために提出されたPCT規則23.1(b)にいう翻訳文の言語
- PCT規則48.3(b)にいう国際公開の言語
- 国際予備審査のために提出されたPCT規則55.2または55.3にいう翻訳文の言語

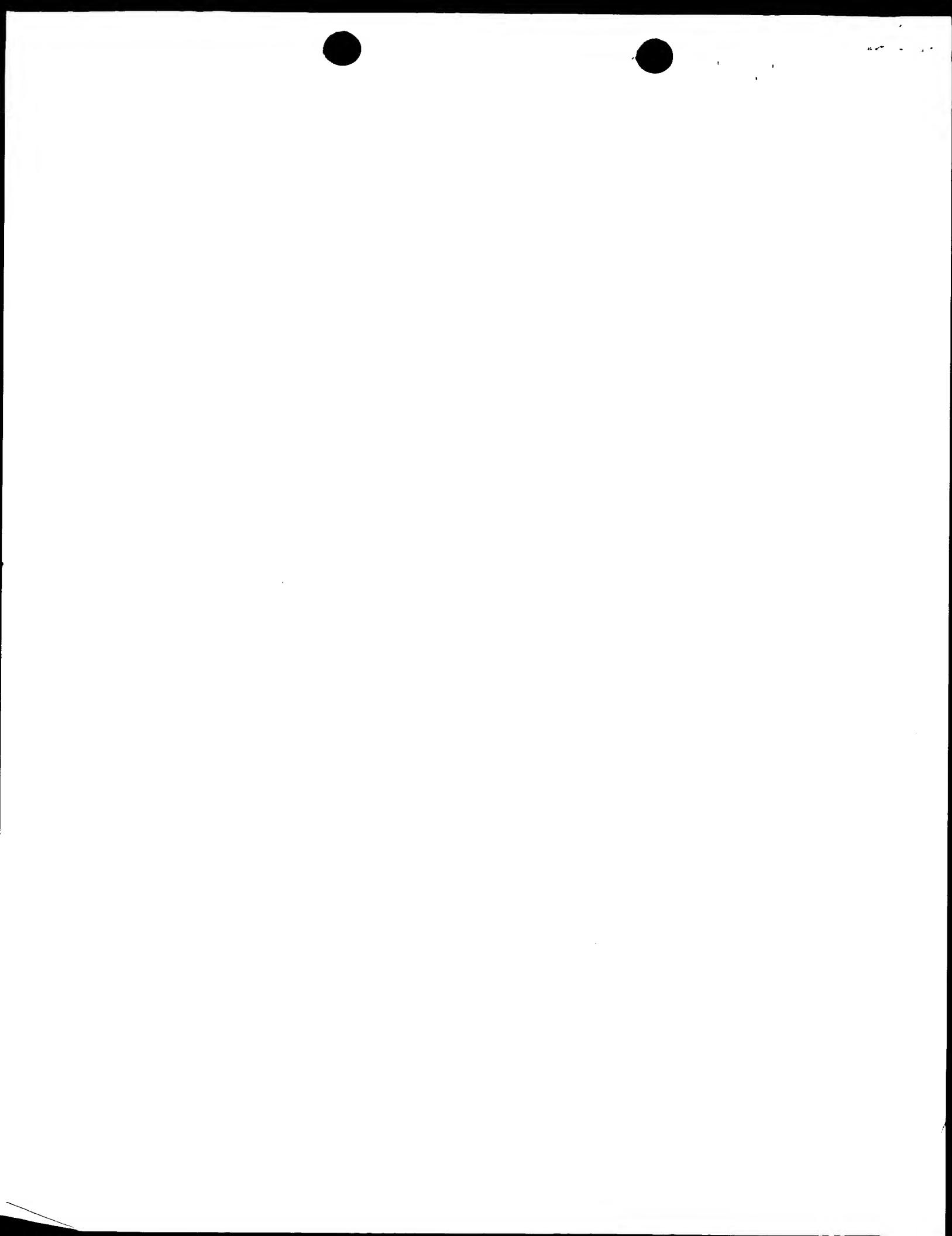
3. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき国際予備審査報告を行った。

- この国際出願に含まれる書面による配列表
- この国際出願と共に提出されたフレキシブルディスクによる配列表
- 出願後に、この国際予備審査（または調査）機関に提出された書面による配列表
- 出願後に、この国際予備審査（または調査）機関に提出されたフレキシブルディスクによる配列表
- 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった
- 書面による配列表に記載した配列とフレキシブルディスクによる配列表に記録した配列が同一である旨の陳述書の提出があった。

4. 補正により、下記の書類が削除された。

- 明細書 第 \_\_\_\_\_ ページ
- 請求の範囲 第 \_\_\_\_\_ 項
- 図面 図面の第 \_\_\_\_\_ ページ/図

5.  この国際予備審査報告は、補充欄に示したように、補正が出願時における開示の範囲を越えてされたものと認められるので、その補正がされなかつたものとして作成した。（PCT規則70.2(c) この補正を含む差し替え用紙は上記1.における判断の際に考慮しなければならず、本報告に添付する。）



## III. 新規性、進歩性又は産業上の利用可能性についての国際予備審査報告の不作成

1. 次に関して、当該請求の範囲に記載されている発明の新規性、進歩性又は産業上の利用可能性につき、次の理由により審査しない。

国際出願全体

請求の範囲 11.

理由：

この国際出願又は請求の範囲 11 は、国際予備審査をすることを要しない次の事項を内容としている（具体的に記載すること）。

データベースは情報の単なる提示に相当する。

明細書、請求の範囲若しくは図面（次に示す部分）又は請求の範囲 の記載が、不明確であるため、見解を示すことができない（具体的に記載すること）。

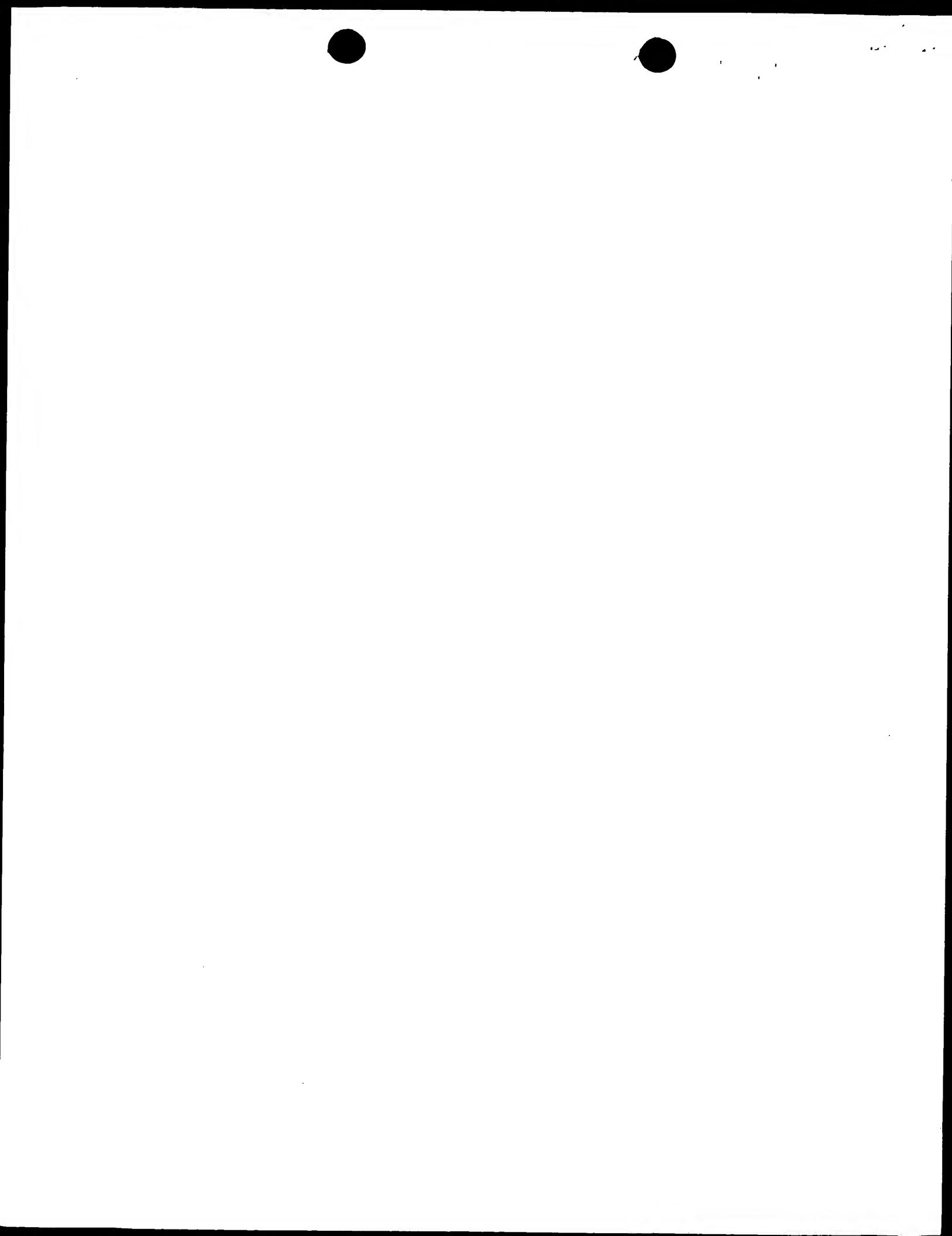
全部の請求の範囲又は請求の範囲 が、明細書による十分な裏付けを欠くため、見解を示すことができない。

請求の範囲 について、国際調査報告が作成されていない。

2. ヌクレオチド又はアミノ酸の配列表が実施細則の附属書C（塩基配列又はアミノ酸配列を含む明細書等の作成のためのガイドライン）に定める基準を満たしていないので、有効な国際予備審査をすることができない。

書面による配列表が提出されていない又は所定の基準を満たしていない。

フレキシブルディスクによる配列表が提出されていない又は所定の基準を満たしていない。



## V. 新規性、進歩性又は産業上の利用可能性についての法第12条 (PCT35条(2)) に定める見解、それを裏付ける文献及び説明

## 1. 見解

新規性 (N) 請求の範囲 1 ~ 10 有  
                  請求の範囲 \_\_\_\_\_ 無

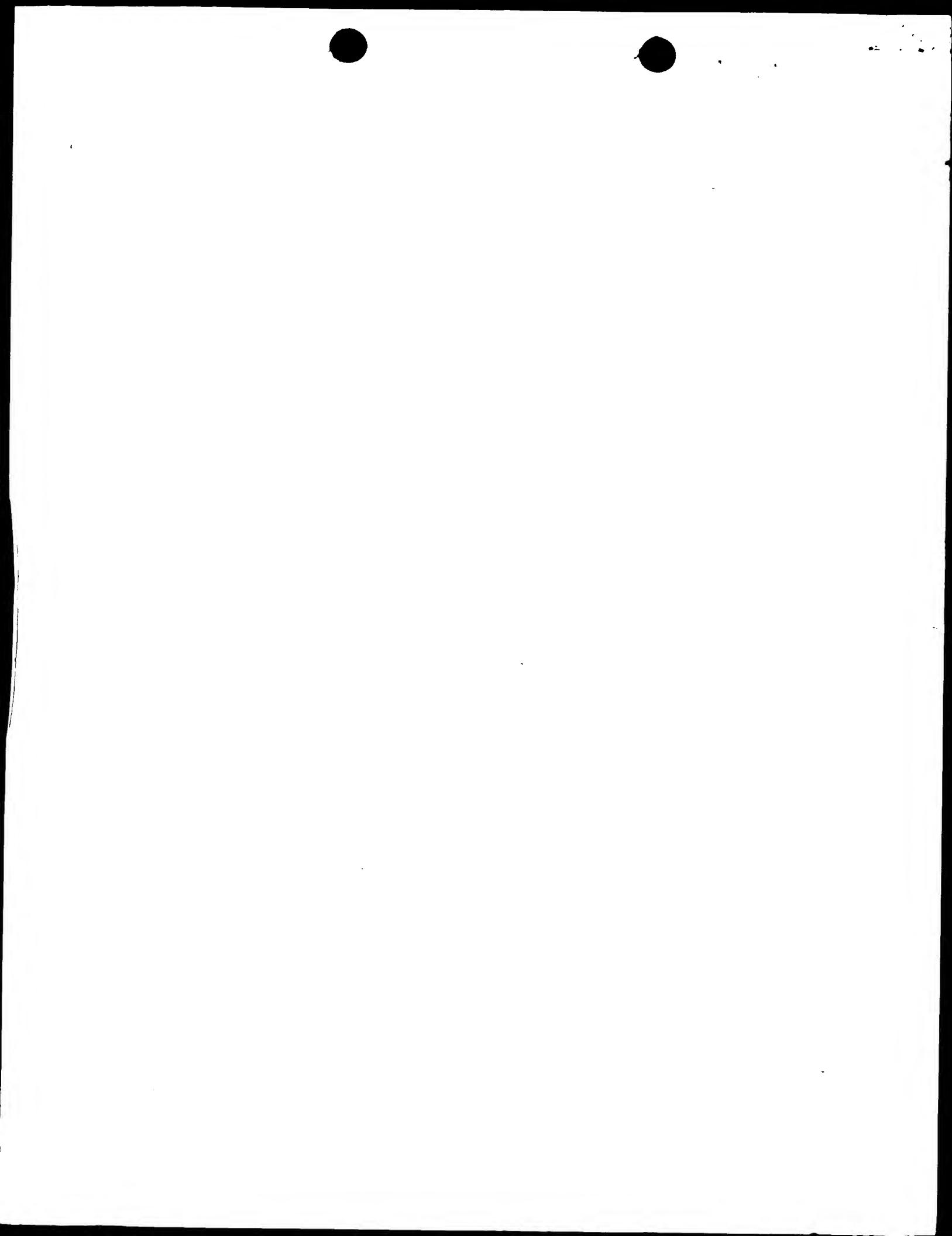
進歩性 (I S) 請求の範囲 \_\_\_\_\_ 有  
                  請求の範囲 1 ~ 10 無

産業上の利用可能性 (I A) 請求の範囲 1 ~ 10 有  
                  請求の範囲 \_\_\_\_\_ 無

## 2. 文献及び説明 (PCT規則70.7)

請求の範囲 1 ~ 10  
文献1: J P, 7-206894, A (アドバンスドテクノロジーインスティチュート株式会社) 8. 8月. 1995 (08. 08. 95)

文献1には、蛋白質の立体構造の予測に際し、立体構造既知の蛋白質で相同性が高いものを選択して検討することが記載されている。その場合に、極性度やpK値等の物理化学的パラメータについて検討することも記載されている。  
パラメータとしてアミノ酸残基の側鎖の疎水性や親水性を用いることは当業者にとって自明である。



# PCT

## REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference  
(if desired) (12 characters maximum) 98215M

### Box No. I TITLE OF INVENTION

METHOD OF INFERRING THREE-DIMENSIONAL STRUCTURE OF PROTEIN

### Box No. II APPLICANT

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

ITAI Akiko

5-16-6, Hongo, Bunkyo-ku, Tokyo  
113-0033 JAPAN

This person is also inventor.

Telephone No.

Facsimile No.

Teleprinter No.

State (that is, country) of nationality: JAPAN

State (that is, country) of residence: JAPAN

This person is applicant for the purposes of:  all designated States  all designated States except the United States of America  the United States of America only  the States indicated in the Supplemental Box

### Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

This person is:

applicant only

applicant and inventor

inventor only (If this check-box is marked, do not fill in below.)

State (that is, country) of nationality:

State (that is, country) of residence:

This person is applicant for the purposes of:  all designated States  all designated States except the United States of America  the United States of America only  the States indicated in the Supplemental Box

Further applicants and/or (further) inventors are indicated on a continuation sheet.

### Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:

agent

common representative

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)

Telephone No.

03-3271-1331

9621 Patent Attorney IMAMURA Masazumi  
9263 Patent Attorney SHIOZAWA Hisao  
9584 Patent Attorney KAMATA Junji

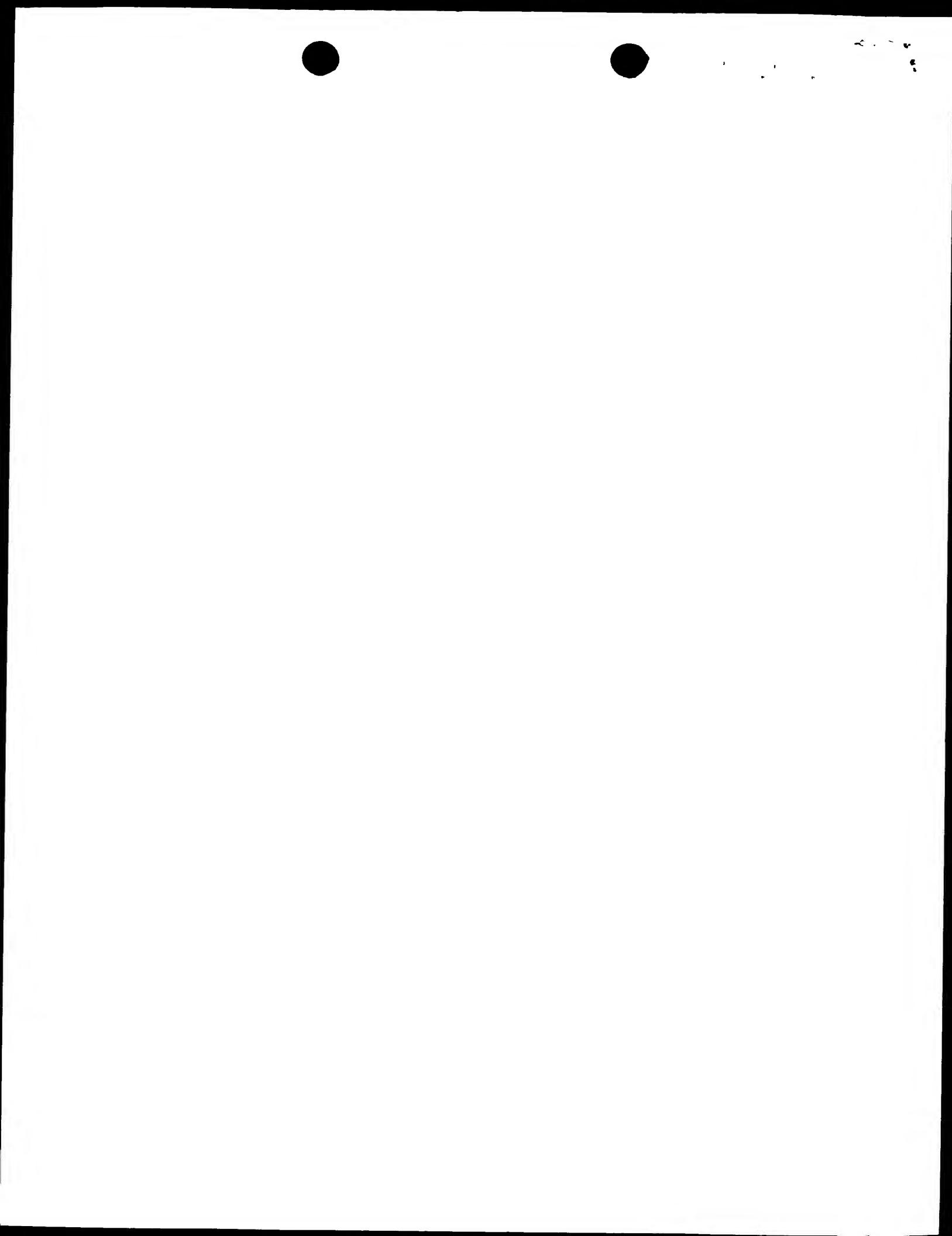
Facsimile No.

03-3271-1410

5th Floor, KRF Bldg., 5-5, Kyobashi 1-chome,  
Chuo-ku, Tokyo 104-0031 JAPAN

Teleprinter No.

Address for correspondence: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.



## Box No.V DESIGNATION OF STATES

The following designations are hereby made under Rule 4.9(a) (mark the applicable check-boxes; at least one must be marked):

## Regional Patent

AP ARIPO Patent: GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, SD Sudan, SZ Swaziland, UG Uganda, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT

EA Eurasian Patent: AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT

EP European Patent: AT Austria, BE Belgium, CH and LI Switzerland and Liechtenstein, CY Cyprus, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT

OA OAPI Patent: BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT (if other kind of protection or treatment desired, specify on dotted line)

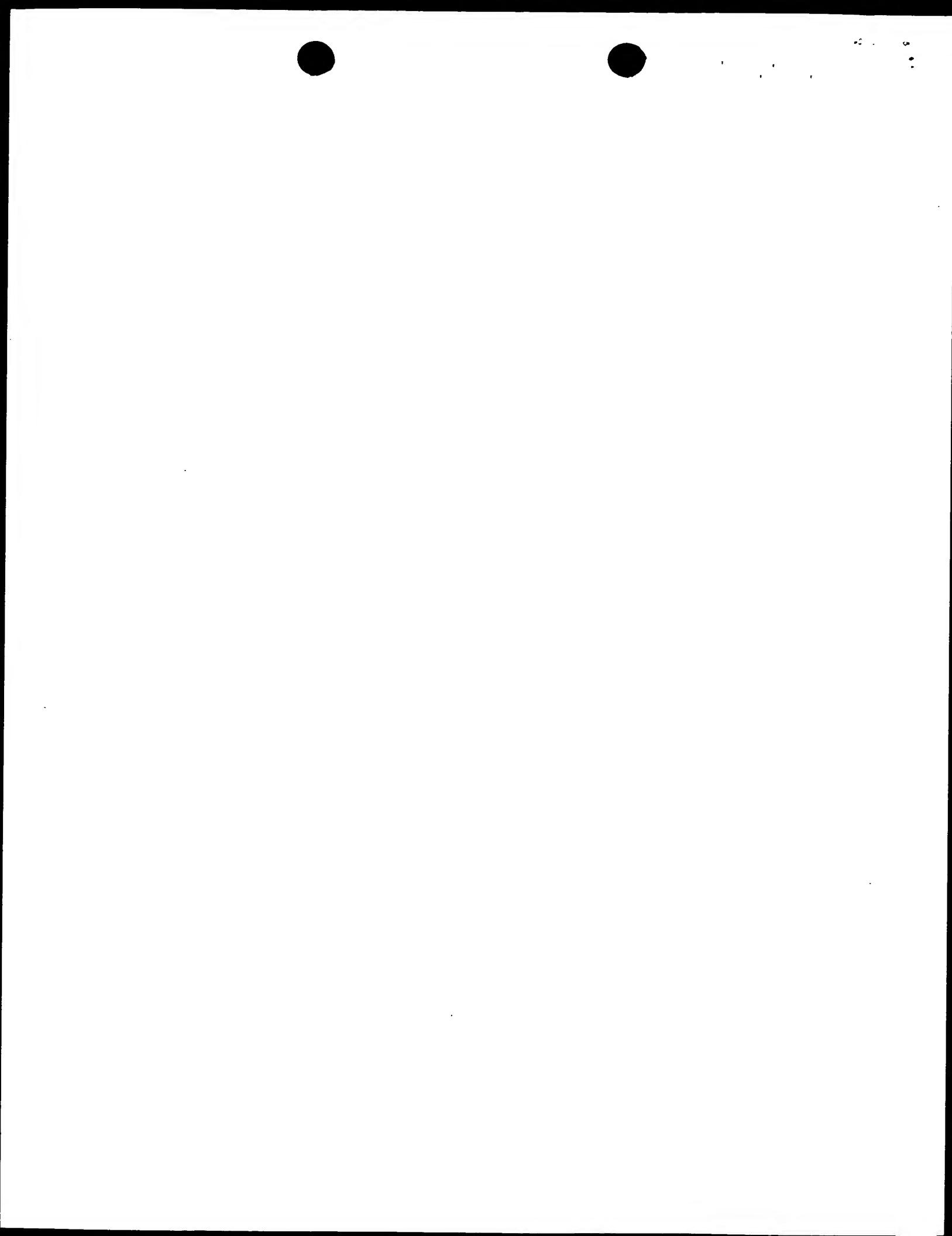
## National Patent (if other kind of protection or treatment desired, specify on dotted line):

<input type="checkbox"/> AL Albania	<input type="checkbox"/> LS Lesotho
<input type="checkbox"/> AM Armenia	<input type="checkbox"/> LT Lithuania
<input type="checkbox"/> AT Austria	<input type="checkbox"/> LU Luxembourg
<input type="checkbox"/> AU Australia	<input type="checkbox"/> LV Latvia
<input type="checkbox"/> AZ Azerbaijan	<input type="checkbox"/> MD Republic of Moldova
<input type="checkbox"/> BA Bosnia and Herzegovina	<input type="checkbox"/> MG Madagascar
<input type="checkbox"/> BB Barbados	<input type="checkbox"/> MK The former Yugoslav Republic of Macedonia
<input type="checkbox"/> BG Bulgaria	<input type="checkbox"/> MN Mongolia
<input type="checkbox"/> BR Brazil	<input type="checkbox"/> MW Malawi
<input type="checkbox"/> BY Belarus	<input type="checkbox"/> MX Mexico
<input type="checkbox"/> CA Canada	<input type="checkbox"/> NO Norway
<input type="checkbox"/> CH and LI Switzerland and Liechtenstein	<input type="checkbox"/> NZ New Zealand
<input type="checkbox"/> CN China	<input type="checkbox"/> PL Poland
<input type="checkbox"/> CU Cuba	<input type="checkbox"/> PT Portugal
<input type="checkbox"/> CZ Czech Republic	<input type="checkbox"/> RO Romania
<input type="checkbox"/> DE Germany	<input type="checkbox"/> RU Russian Federation
<input type="checkbox"/> DK Denmark	<input type="checkbox"/> SD Sudan
<input type="checkbox"/> EE Estonia	<input type="checkbox"/> SE Sweden
<input type="checkbox"/> ES Spain	<input type="checkbox"/> SG Singapore
<input type="checkbox"/> FI Finland	<input type="checkbox"/> SI Slovenia
<input type="checkbox"/> GB United Kingdom	<input type="checkbox"/> SK Slovakia
<input type="checkbox"/> GE Georgia	<input type="checkbox"/> SL Sierra Leone
<input type="checkbox"/> GH Ghana	<input type="checkbox"/> TJ Tajikistan
<input type="checkbox"/> GM Gambia	<input type="checkbox"/> TM Turkmenistan
<input type="checkbox"/> GW Guinea-Bissau	<input type="checkbox"/> TR Turkey
<input type="checkbox"/> HR Croatia	<input type="checkbox"/> TT Trinidad and Tobago
<input type="checkbox"/> HU Hungary	<input type="checkbox"/> UA Ukraine
<input type="checkbox"/> ID Indonesia	<input type="checkbox"/> UG Uganda
<input type="checkbox"/> IL Israel	<input checked="" type="checkbox"/> US United States of America
<input type="checkbox"/> IS Iceland	<input type="checkbox"/> UZ Uzbekistan
<input checked="" type="checkbox"/> JP Japan	<input type="checkbox"/> VN Viet Nam
<input type="checkbox"/> KE Kenya	<input type="checkbox"/> YU Yugoslavia
<input type="checkbox"/> KG Kyrgyzstan	<input type="checkbox"/> ZW Zimbabwe
<input type="checkbox"/> KP Democratic People's Republic of Korea	
<input type="checkbox"/> KR Republic of Korea	
<input type="checkbox"/> KZ Kazakhstan	
<input type="checkbox"/> LC Saint Lucia	
<input type="checkbox"/> LK Sri Lanka	
<input type="checkbox"/> LR Liberia	

Check-boxes reserved for designating States (for the purposes of a national patent) which have become party to the PCT after issuance of this sheet:

.....  
 .....

**Precautionary Designation Statement:** In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. (Confirmation of a designation consists of the filing of a notice specifying that designation and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.)



## Box No. VI 'PRIORITY CLAIM'

 Further priority claims are indicated in the Supplemental Box.

Filing date of earlier application (day/month/year)	Number of earlier application	Where earlier application is:		
		national application: country	regional application: regional Office	international application: receiving Office
item (1) 02/10/97	Patent Application 9-269611	JAPAN		
item (2)				
item (3)				

The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) (only if the earlier application was filed with the Office which for the purposes of the present international application is the receiving Office) identified above as item(s):

(1)

\* Where the earlier application is an ARIPO application, it is mandatory to indicate in the Supplemental Box at least one country party to the Paris Convention for the Protection of Industrial Property for which that earlier application was filed (Rule 4.10(b)(ii)). See Supplemental Box.

## Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA)  
(if two or more International Searching Authorities are  
competent to carry out the international search, indicate  
the Authority chosen; the two-letter code may be used):

ISA / JP

Request to use results of earlier search; reference to that search (if an earlier  
search has been carried out by or requested from the International Searching Authority):

Date (day/month/year)

Number

Country (or regional Office)

## Box No. VIII CHECK LIST; LANGUAGE OF FILING

This international application contains  
the following number of sheets:

request	:	3
description (excluding sequence listing part)	:	18
claims	:	2
abstract	:	1
drawings	:	1
sequence listing part of description	:	0
Total number of sheets	:	25

This international application is accompanied by the item(s) marked below:

- fee calculation sheet
- separate signed power of attorney
- copy of general power of attorney; reference number, if any:
- statement explaining lack of signature
- priority document(s) identified in Box No. VI as item(s):
- translation of international application into (language):
- separate indications concerning deposited microorganism or other biological material
- nucleotide and/or amino acid sequence listing in computer readable form
- other (specify): Request for transmittal of priority document

Figure of the drawings which  
should accompany the abstract:

Language of filing of the  
international application: Japanese

## Box No. IX SIGNATURE OF APPLICANT OR AGENT

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

IMAMURA Masazumi

SHIOZAWA Hisao

KAMATA Junji

## For receiving Office use only

1. Date of actual receipt of the purported  
international application:

3. Corrected date of actual receipt due to later but  
timely received papers or drawings completing  
the purported international application:

4. Date of timely receipt of the required  
corrections under PCT Article 11(2):

5. International Searching Authority  
(if two or more are competent): ISA / JP

2. Drawings:

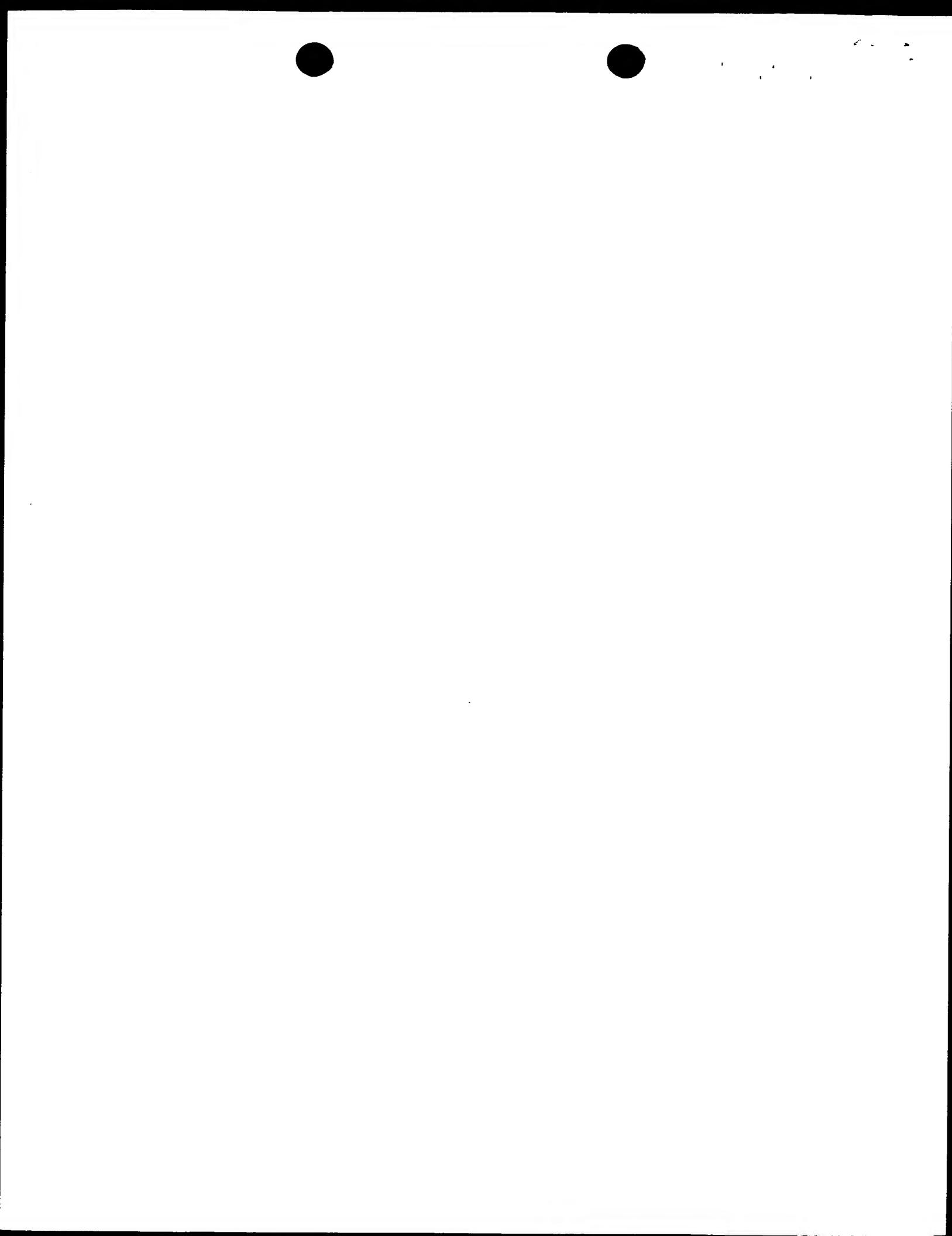
received:

not received:

6.  Transmittal of search copy delayed  
until search fee is paid.

## For International Bureau use only

Date of receipt of the record copy  
by the International Bureau:



This sheet is not part of and does not count as a sheet of the international application.

# PCT

## FEE CALCULATION SHEET Annex to the Request

For receiving Office use only

International application No.

Date stamp of the receiving Office

Applicant's or agent's  
file reference 98215M

Applicant  
ITAI Akiko

### CALCULATION OF PRESCRIBED FEES

1. TRANSMITTAL FEE . . . . . 95,000 (T+S)  T

2. SEARCH FEE . . . . .  S

International search to be carried out by \_\_\_\_\_  
(If two or more International Searching Authorities are competent in relation to the international application, indicate the name of the Authority which is chosen to carry out the international search.)

3. INTERNATIONAL FEE

#### Basic Fee

The international application contains 25 sheets.

first 30 sheets . . . . . 55,000  b1

0 x 1,300 = 0  b2

remaining sheets additional amount

Add amounts entered at b1 and b2 and enter total at B . . . . . 55,000  B

#### Designation Fees

The international application contains 3 designations.

3 x 12,700 = 38,100  D

number of designation fees amount of designation fee  
payable (maximum 11)

Add amounts entered at B and D and enter total at I . . . . . 93,100  I

(Applicants from certain States are entitled to a reduction of 75% of the international fee. Where the applicant is (or all applicants are) so entitled, the total to be entered at I is 25% of the sum of the amounts entered at B and D.)

4. FEE FOR PRIORITY DOCUMENT (if applicable) . . . . .  P

5. TOTAL FEES PAYABLE . . . . .  188,100

Add amounts entered at T, S, I and P, and enter total in the TOTAL box

**TOTAL**

The designation fees are not paid at this time.

### MODE OF PAYMENT

authorization to charge  
deposit account (see below)  
 cheque  
 postal money order

bank draft  
 cash  
 revenue stamps

coupons  
 other (specify): \_\_\_\_\_

### DEPOSIT ACCOUNT AUTHORIZATION (this mode of payment may not be available at all receiving Offices)

The RO/  is hereby authorized to charge the total fees indicated above to my deposit account.

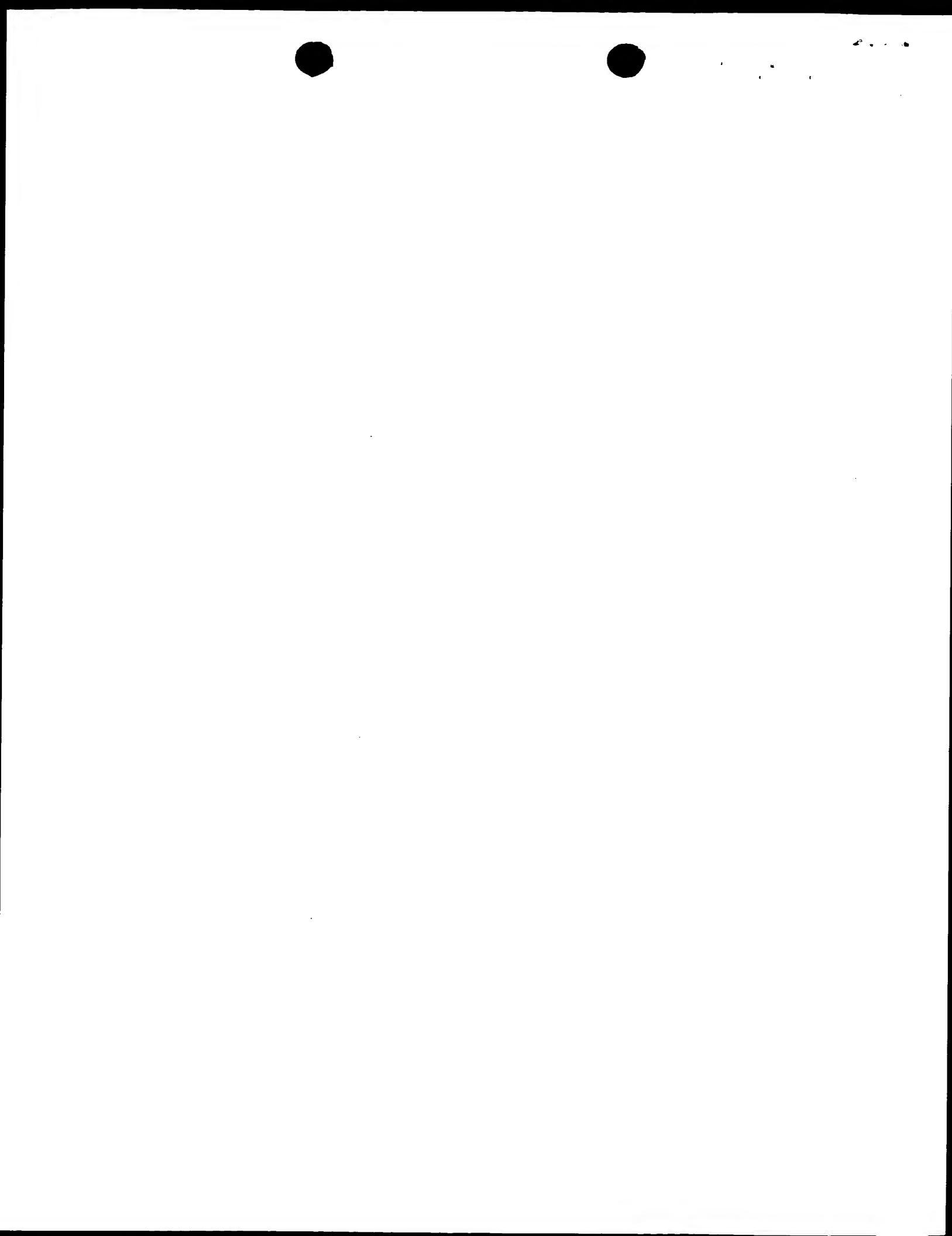
is hereby authorized to charge any deficiency or credit any overpayment in the total fees indicated above to my deposit account.

is hereby authorized to charge the fee for preparation and transmittal of the priority document to the International Bureau of WIPO to my deposit account.

Deposit Account No.

Date (day/month/year)

Signature



50003  
091508527  
Translation

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference 98215M	<b>FOR FURTHER ACTION</b> See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/JP98/04457	International filing date (day/month/year) 02 October 1998 (02.10.1998)	Priority date (day/month/year) 02 October 1997 (02.10.1997)
International Patent Classification (IPC) or national classification and IPC G01N 33/68, G06F 17/00, 15/18, C07K 14/00		
Applicant ITAI, Akiko		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 4 sheets, including this cover sheet.

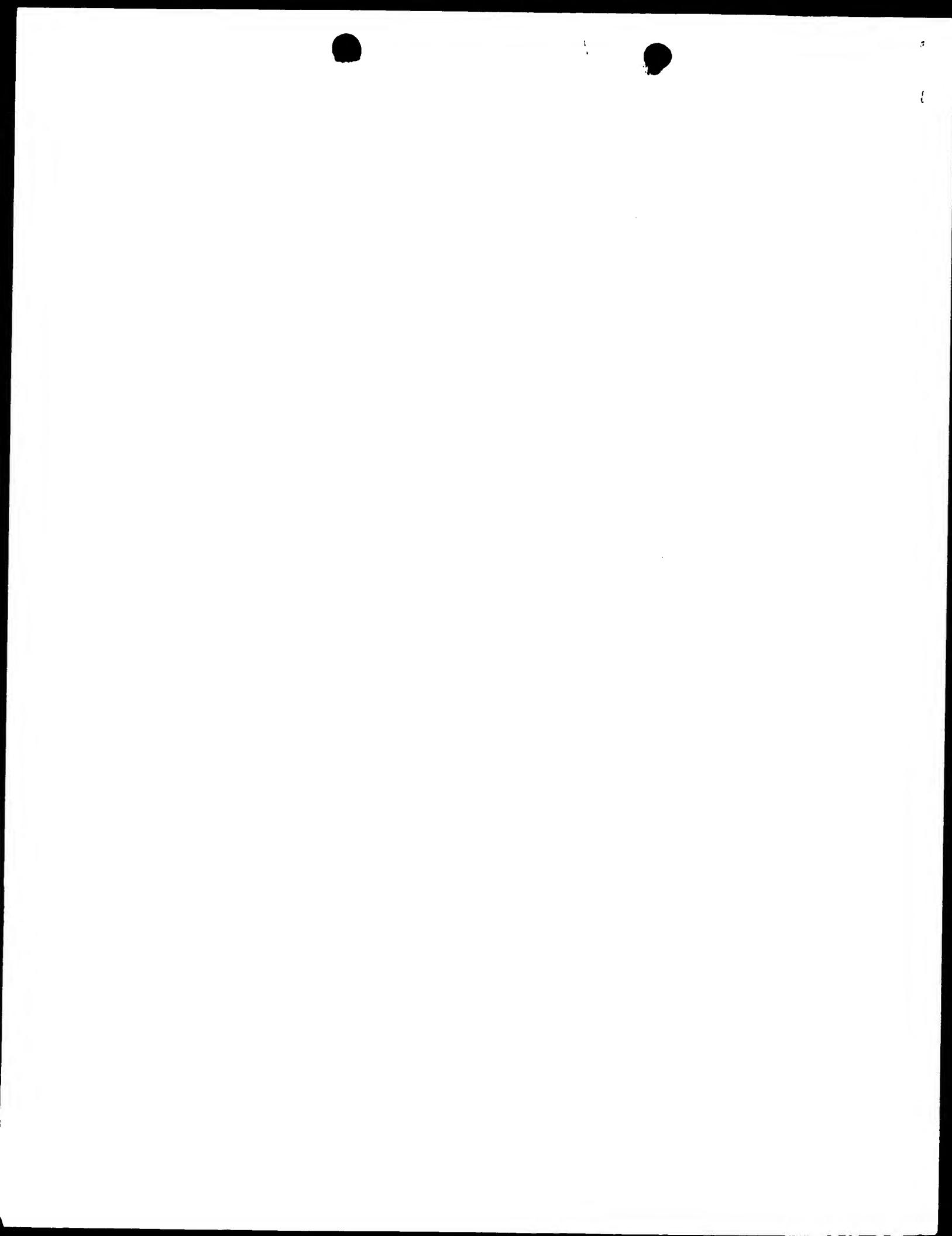
This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of \_\_\_\_\_ sheets.

3. This report contains indications relating to the following items:

- I  Basis of the report
- II  Priority
- III  Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV  Lack of unity of invention
- V  Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI  Certain documents cited
- VII  Certain defects in the international application
- VIII  Certain observations on the international application

Date of submission of the demand 02 October 1998 (02.10.1998)	Date of completion of this report 30 April 1999 (30.04.1999)
Name and mailing address of the IPEA/JP Japanese Patent Office, 4-3 Kasumigaseki 3-chome Chiyoda-ku, Tokyo 100-8915, Japan Facsimile No.	Authorized officer Telephone No. (81-3) 3581 1101



## INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/JP98/04457

## I. Basis of the report

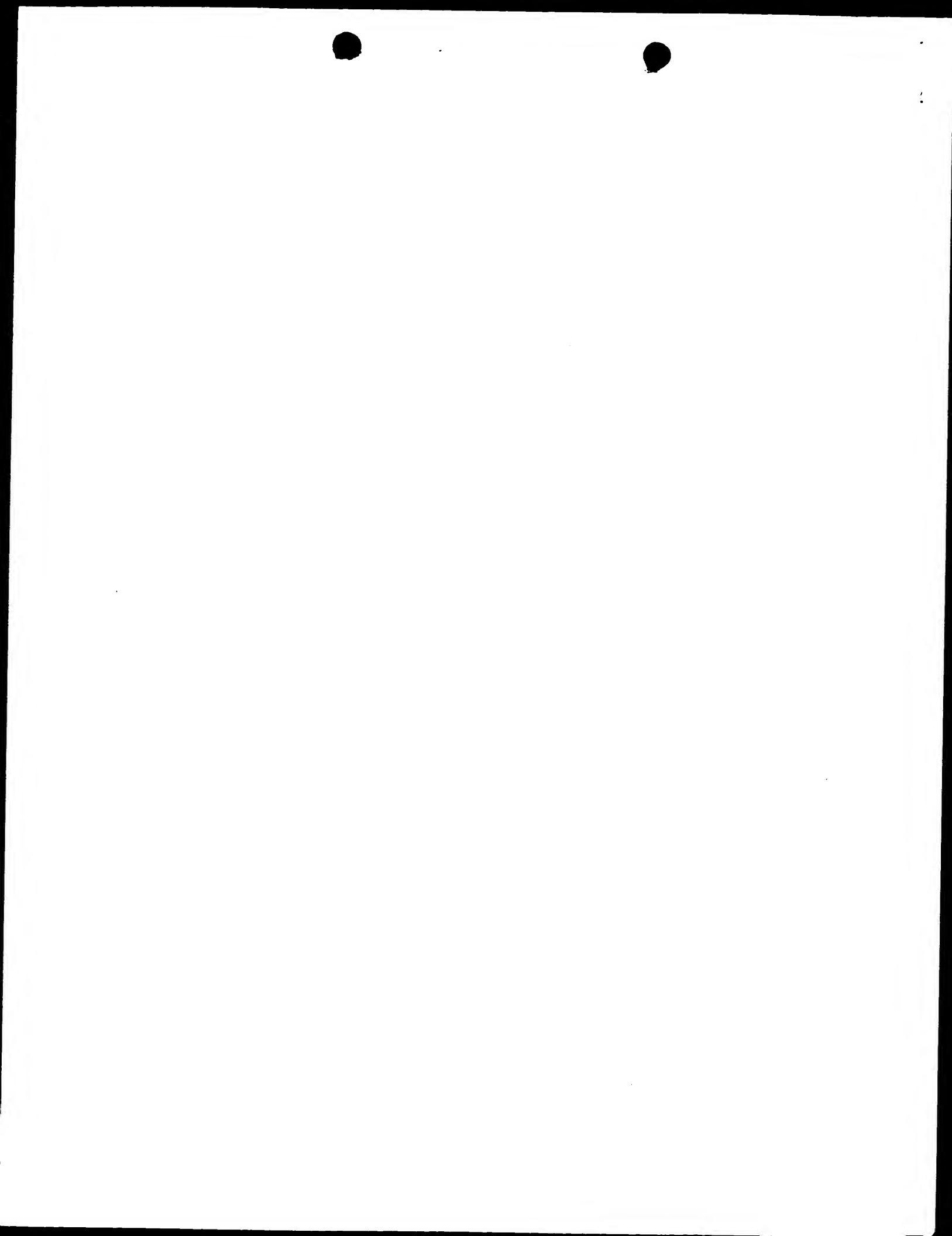
1. With regard to the **elements** of the international application:\* the international application as originally filed the description:pages \_\_\_\_\_, as originally filed  
pages \_\_\_\_\_, filed with the demand  
pages \_\_\_\_\_, filed with the letter of \_\_\_\_\_ the claims:pages \_\_\_\_\_, as originally filed  
pages \_\_\_\_\_, as amended (together with any statement under Article 19)  
pages \_\_\_\_\_, filed with the demand  
pages \_\_\_\_\_, filed with the letter of \_\_\_\_\_ the drawings:pages \_\_\_\_\_, as originally filed  
pages \_\_\_\_\_, filed with the demand  
pages \_\_\_\_\_, filed with the letter of \_\_\_\_\_ the sequence listing part of the description:pages \_\_\_\_\_, as originally filed  
pages \_\_\_\_\_, filed with the demand  
pages \_\_\_\_\_, filed with the letter of \_\_\_\_\_2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language \_\_\_\_\_ which is:

 the language of a translation furnished for the purposes of international search (under Rule 23.1(b)). the language of publication of the international application (under Rule 48.3(b)). the language of the translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing: contained in the international application in written form. filed together with the international application in computer readable form. furnished subsequently to this Authority in written form. furnished subsequently to this Authority in computer readable form. The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished. The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.4.  The amendments have resulted in the cancellation of: the description, pages \_\_\_\_\_ the claims, Nos. \_\_\_\_\_ the drawings, sheets/fig \_\_\_\_\_5.  This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).\*\*

\* Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rule 70.16 and 70.17).

\*\* Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.



## III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non obvious), or to be industrially applicable have not been examined in respect of:

the entire international application.

claims Nos. 11

because:

the said international application, or the said claims Nos. 11 relate to the following subject matter which does not require an international preliminary examination (specify):

A database corresponds to the mere presentation of information.

the description, claims or drawings (*indicate particular elements below*) or said claims Nos. \_\_\_\_\_ are so unclear that no meaningful opinion could be formed (specify):

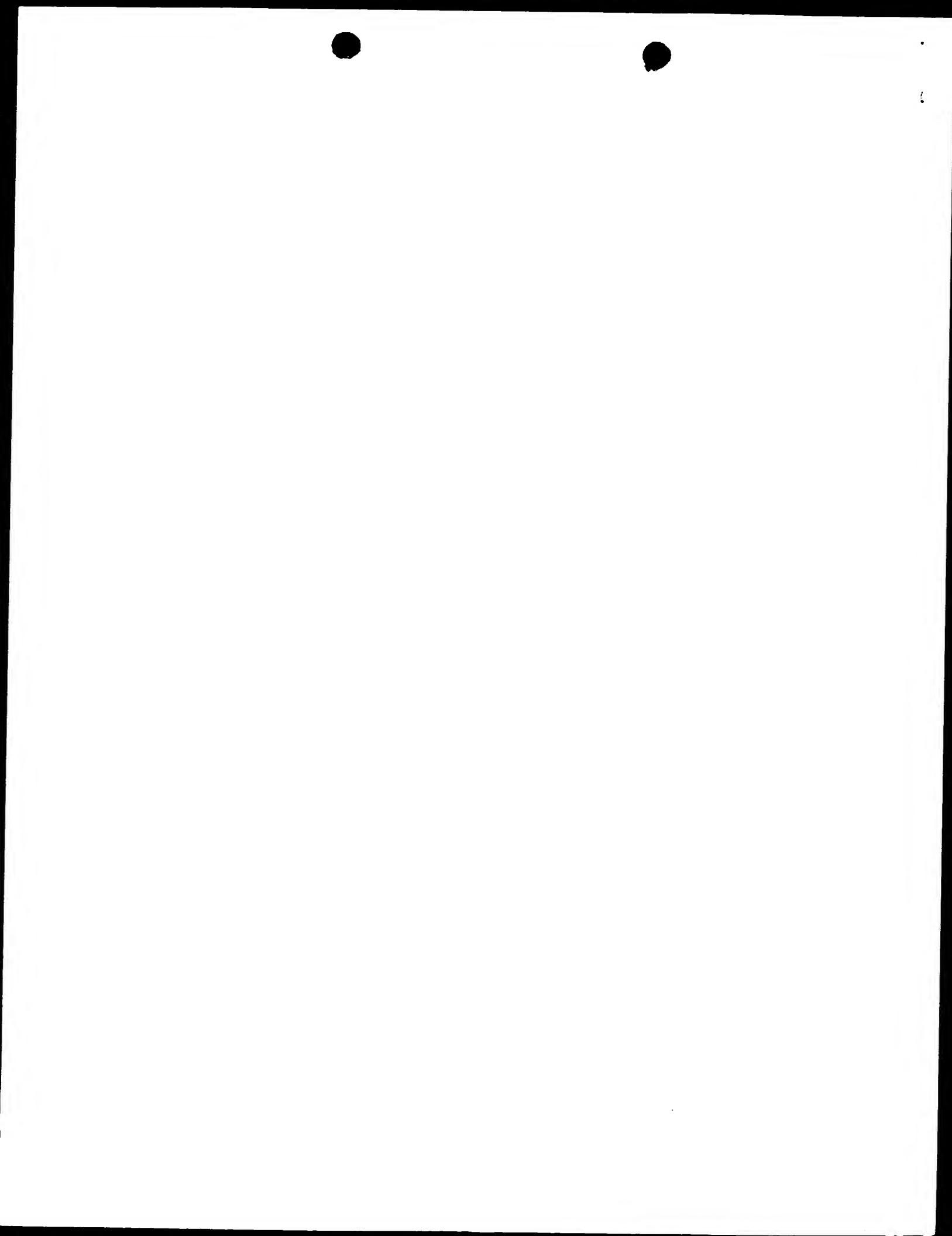
the claims, or said claims Nos. \_\_\_\_\_ are so inadequately supported by the description that no meaningful opinion could be formed.

no international search report has been established for said claims Nos. \_\_\_\_\_

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

the written form has not been furnished or does not comply with the standard.

the computer readable form has not been furnished or does not comply with the standard.



## INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/JP98/04457

## V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

## 1. Statement

Novelty (N)	Claims	1-10	YES
	Claims		NO
Inventive step (IS)	Claims		YES
	Claims	1-10	NO
Industrial applicability (IA)	Claims	1-10	YES
	Claims		NO

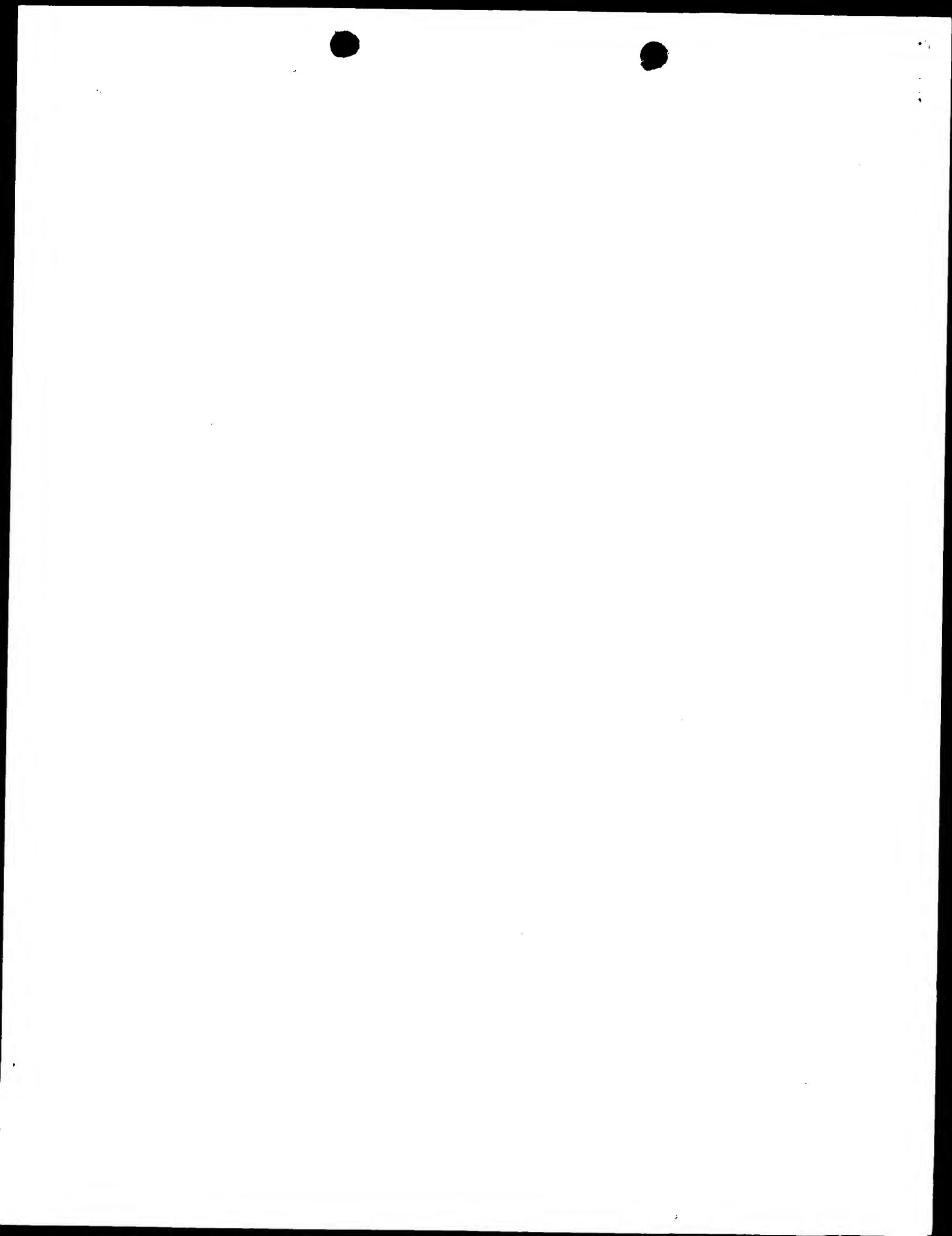
## 2. Citations and explanations

## Claims 1-10

Document 1: JP, 7-206894, A (Advanced Technology Institute K.K.), 8 August, 1995 (08.08.95)

Document 1 discloses the idea, when predicting the three-dimensional structure of a protein, of carrying out investigations whereby proteins that have a known three-dimensional structure and are highly homologous to the protein of unknown three-dimensional structure are selected. It also discloses the idea of investigating physico-chemical parameters such as degree of polarity and pK value in this case.

It is considered that it would be obvious to a person skilled in the art to use the hydrophobicity or hydrophilicity of the side chains of the amino acid residues as such a parameter.



A. 発明の属する分野の分類 (国際特許分類 (IPC))  
Int. Cl. G01N33/68, G06F17/00, G06F15/18, C07K14/00

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))  
Int. Cl. G01N33/68, G06F17/00, G06F15/18, C07K14/00

## 最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-1998年
日本国登録実用新案公報	1994-1998年
日本国実用新案登録公報	1996-1998年

## 国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
PA	JP, 10-95796, A (日本電気株式会社) 14. 4月. 1998 (14. 04. 98) (ファミリーなし)	1~10
X	JP, 7-206894, A (アドバンスドテクノロジーインスティチュート株式会社) 8. 8月. 1995 (08. 08. 95) (ファミリーなし)	1~10
A	JP, 8-69446, A (日本電気株式会社) 12. 3月. 1996 (12. 03. 96) (ファミリーなし)	1~10
A	JP, 7-105179, A (日本電気株式会社) 21. 4月. 1995 (21. 04. 95) (ファミリーなし)	1~10

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの  
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
「O」口頭による開示、使用、展示等に言及する文献  
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

## の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
「&」同一パテントファミリー文献

## 国際調査を完了した日

14. 12. 98-

## 国際調査報告の発送日

06.01.99

## 国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

## 特許庁審査官 (権限のある職員)

亀田 宏之

2 J 9015

電話番号 03-3581-1101 内線 3252

C (続き) . 関連すると認められる文献		関連する 請求の範囲の番号
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	
A	JP, 7-93286, A (日本電気株式会社) 7. 4月. 1995 (07. 04. 95) (ファミリーなし)	1~10
A	JP, 7-13959, A (日本電気株式会社) 17. 1月. 1995 (17. 01. 95) (ファミリーなし)	1~10
A	JP, 5-282383, A (富士通株式会社) 29. 10月. 1993 (29. 10. 93) (ファミリーなし)	1~10
A	JP, 4-45781, A (富士通株式会社) 14. 2月. 1992 (14. 02. 92) (ファミリーなし)	1~10

PCT

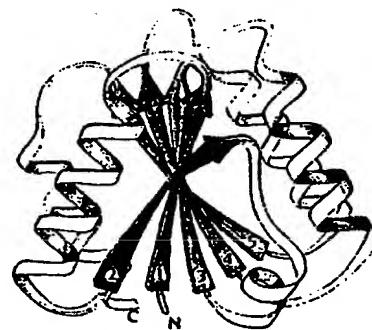
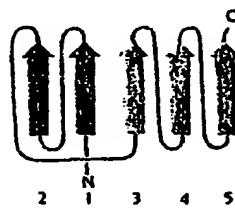
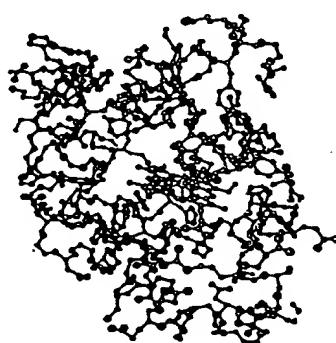
世界知的所有権機関  
国際事務局  
特許協力条約に基づいて公開された国際出願



(51) 国際特許分類6 G01N 33/68, G06F 17/00, 15/18, C07K 14/00		A1	(11) 国際公開番号 WO99/18440
(21) 国際出願番号 PCT/JP98/04457			(43) 国際公開日 1999年4月15日(15.04.99)
(22) 国際出願日 1998年10月2日(02.10.98)			(81) 指定国 JP, US, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
(30) 優先権データ 特願平9/269611	1997年10月2日(02.10.97)	JP	添付公開書類 国際調査報告書
(71) 出願人 ; および 板井昭子(ITAL, Akiko)[JP/JP] 〒113-0033 東京都文京区本郷5-16-6 Tokyo, (JP)			(74) 代理人 弁理士 今村正純, 外(IMAMURA, Masazumi et al.) 〒104-0031 東京都中央区京橋1丁目5番5号 KRFビル5階 Tokyo, (JP)

(54)Title: METHOD OF INFERRING THREE-DIMENSIONAL STRUCTURE OF PROTEIN

(54)発明の名称 蛋白質の立体構造の推定方法



## (57) Abstract

A method of inferring a three-dimensional structure of a protein, which comprises cleaving the amino acid sequence of a reference protein having a known or inferable structure into a core part sequence which substantially participates in hydrophobic core formation and a sub-part sequence which does not participate in the formation, employing a database containing environmental information concerning the side chains of the respective amino acid residues to effect matching for respective part sequences on the basis of the environmental information concerning the respective amino acid residues of the reference protein and either the hydrophobicity or the hydrophilicity of the side chains of the respective amino acid residues in a query sequence, and selecting a template protein having a three-dimensional structure which is highly similar to that of the query sequence among a number of the reference proteins, thus inferring the scaffold of the proteins of the query sequence.

立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列を疎水コアの形成に実質的に関与するコア部分配列と関与しないサブ部分配列に分割し、各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、部分配列毎に参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキヤップオールドを推定する方法。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

A E アラブ首長国連邦	E S スペイン	L I リヒテンシュタイン	S G シンガポール
A L アルバニア	F I フィンランド	L K スリ・ランカ	S I スロヴェニア
A M アルメニア	F R フランス	L R リベリア	S K スロヴァキア
A T オーストリア	G A ガボン	L S レソト	S L シエラ・レオネ
A U オーストラリア	G B 英国	L T リトアニア	S N セネガル
A Z アゼルバイジャン	G D グレナダ	L U ルクセンブルグ	S Z スワジ蘭ド
B A ボスニア・ヘルツェゴビナ	G E グルジア	L V ラトヴィア	T D チャード
B B バルバドス	G H ガーナ	MC モナコ	T G トーゴー
B E ベルギー	G M ガンビア	MD モルドヴァ	T J タジキスタン
B F ブルガリア・ファソ	G N ギニア	MG マダガスカル	T M トルクメニスタン
B G ブルガリア	G W ギニア・ビサオ	MK マケドニア旧ユーゴスラヴィア	T R トルコ
B J ベナン	G R ギリシャ	共和国	T T トリニダード・トバゴ
B R ブラジル	H R クロアチア	ML マリ	U A ウクライナ
B Y ベラルーシ	H U ハンガリー	MN モンゴル	U G ウガンダ
C A カナダ	I D インドネシア	MR モーリタニア	U S 米国
C F 中央アフリカ	I E アイルランド	MW マラウイ	U Z ウズベキスタン
C G コンゴー	I L イスラエル	MX モキシコ	V N ヴィエトナム
C H スイス	I N インド	NE ニジエール	Y U ユーゴースラビア
C I コートジボアール	I S アイスランド	N L オランダ	Z A 南アフリカ共和国
C M カメルーン	I T イタリア	NO ノールウェー	Z W ジンバブエ
C N 中国	J P 日本	N Z ニュー・ジーランド	
C U キューバ	KE ケニア	P L ポーランド	
C Y キプロス	K G キルギスタン	P T ポルトガル	
C Z チェコ	K P 北朝鮮	RO ルーマニア	
D E ドイツ	K R 韓国	R U ロシア	
D K デンマーク	K Z カザフスタン	S D スーダン	
E E エストニア	L C セントルシア	S E スウェーデン	

## 明細書

## 蛋白質の立体構造の推定方法

## 技術分野

本発明は、蛋白質の立体構造の推定方法に関するものである。

## 背景技術

アミノ酸配列から蛋白質の立体構造を推定することは理論的に不可能とはいえない。しかしながら、現在のところ、配列情報から蛋白質の立体構造を確実に推定する手段は開発されておらず、蛋白質の立体構造を知るための手段は、X線結晶構造解析やnmr解析などの実験的方法に限られている。蛋白質の立体構造の情報は、その機能を原子レベルで理解し、またその蛋白質を標的とする医薬の創製やさらに優れた機能をもつ有用な蛋白質の創製に不可欠である。近年、遺伝子情報の解析手段が急速に進歩した結果、実際に蛋白質が単離されないまま配列情報だけが解明される例が急増している。従って、配列情報から蛋白質の立体構造や機能を推定する有効な手段の開発が切望されているのが現状である。

あるアミノ酸配列を有する蛋白質の存在がわかった場合、配列データベースから相同性のある蛋白質を検索するのが一般的である。アミノ酸配列の一致度がある程度よい蛋白質が見つかると、さらにその蛋白質との間で相同性やギャップも考慮したアラインメントを行い、さらに相同性の高いアラインメントの探索が行われる。目的の蛋白質と機能既知の蛋白質との相同性が高い場合には機能がその既知蛋白質に類似しており、一方、立体構造既知の蛋白質との相同性が高い場合には立体構造がその蛋白質に類似しているとの推定が成り立つ。また、相同性が高ければ高いほど機能や立体構造が類似している確率が高く、その推定の信頼性が高いと考えられる。

立体構造既知の蛋白質の配列とある程度（一般に30%程度）以上の相同性が

認められた場合には、その立体構造を鋳型として立体構造を構築するホモロジーモデリング法が行われる。鋳型の立体構造に基づいて対応付けられた残基が鋳型と異なる場合には、側鎖を置換することによって仮想的な立体構造を構築することができる。アラインメント上のギャップは、鋳型立体構造中に対応するアミノ酸がないか、あるいは鋳型の方が余計なアミノ酸をもつことになるので別途処理する必要があるが、ギャップの存在は鋳型に基づくモデリング作業を難しくし、かつ信頼性を損なうので、できるだけギャップの数を少なくするようにギャップにペナルティを与えるアラインメント法が推奨されている。

問題のアミノ酸配列とある程度以上相同性の高いアミノ酸配列を有する立体構造既知の蛋白質が見つからない場合には、ホモロジーモデリングは不可能である。一方で、蛋白質の結晶構造情報が蓄積されるにつれて、アミノ酸配列の相同性が殆どなく機能も全く異なる複数の蛋白質が、類似の立体構造を持つ例が多数わかつてきた。このことは、蛋白質が安定な立体構造をとるための物理化学的要因を考慮すれば、アミノ酸配列の相同性が低い場合にも、立体構造既知の蛋白質群中から鋳型となる立体構造を見出せる可能性があることを示している。

近年、アミノ酸残基毎の疎水性など物理的な性質の一致を考慮したスコアを用いることにより、アミノ酸配列の相同性が低くても立体構造的に類似性が高いと推定される鋳型蛋白質を立体構造既知の蛋白質群から選ぶ方法が開発された。代表的な方法として、Eisenberg らによる 3D-1D 法 (R. Luthy, J. U. Bowie and D. Eisenberg, *Nature*, 356, 83, 1992) がある。この方法は、アミノ酸配列の相同性に加え、立体構造既知の蛋白質について各アミノ酸残基の属する 2 次構造とその残基位置の環境を表すパラメータと、問題の配列の各アミノ酸残基に対して 2 次構造毎に与えたパラメータとを用いて、対応づけられたアミノ酸残基間で類似性のスコアを計算する工程を含んでいる。この方法では、蛋白質のペプチド鎖の折り畳み方についての膨大な自由度の問題を既知の結晶構造群を鋳型として用いることで回避しており、評価の要素として疎水性などの物理的パラメータを含めることによって配列の相同性が低い場合にもモデリングが可能である。

しかしながら、立体構造が類似している場合であっても、アミノ酸残基数、2次構造、及びループの長さが同じことは少ないので、アミノ酸配列間の対応づけに基づいた 3D-1D 法を実際に適用する場合には数多くの問題が生じる。例えば、アミノ酸配列間の単純なスライド（スレッディング）に加え、どちらかの配列に部分的な配列の欠損など（ギャップ）を考慮して対応付けする必要があるが、ホモロジーモデリングと同様にギャップの導入はモデリングの信頼性を低下させるてしまう。配列の相同性が低いときに、必要にして最小限のギャップを考慮した対応付けをいかにして実現するかが問題である。また、疎水性親水性などのパラメータの他、20種類のアミノ酸残基に対して2次構造毎のパラメータを用いるなど非常に多数のパラメータに依存しており、パラメータの改良による予測性の向上は期待しにくい。

アミノ酸配列から蛋白質の立体構造を予測する研究の歴史は、配列のどの部分がどのような2次構造をとるかを予測することから始まった。多数の蛋白質の結晶構造情報から統計的に求められたアミノ酸残基毎あるいは連続した数アミノ酸残基の組毎に、 $\alpha$ -ヘリックス又は $\beta$ -シートになり易さを示すパラメータを用いて、問題のアミノ酸配列から顕著な傾向を示す連続した領域を検出し、それがどちらの構造をとるかを予測するものである。その代表的なものとして、Chou と Fasman らによる2次構造予測法 (P. Y. Chou, & G. D. Fasman, Adv. enzymol. 47, 45, 1978) がある。しかしながら、このような方法は、2次構造の3次元的集合に関してはなんら情報を与えるものではなく、また予測された2次構造と結晶解析で確認された2次構造の一致が 60% 前後であることから、立体構造の推定方法としてはほとんど利用価値がない。

蛋白質の安定な折り畳み構造を、純粹に計算的手法によって先入観を入れずに予測する試みも行われるようになった（いわゆる *ab initio* 予測法）。しかしながら、蛋白質は極めて自由度が大きい分子であり（100 残基程度の蛋白質でも考慮すべき自由度のパラメータは 400 以上である）、すべての自由度を考慮して可能な構造を十分に探索することは、現在利用可能なコンピュータでは不可能で

ある。また、可能な立体構造の安定性を正しく評価できるほど蛋白質の構造の安定化に関わるファクター（例えば、水の物理化学的性質、疎水相互作用、静電相互作用）についての研究は進んでいないなどの理由から、このような構造予測法の成功は今のところ期待できない。

一方、近年、多数の蛋白質の立体構造が解析されており、その成果はプロテインデータバンクから利用できるようになっている。現在約 6,000 の蛋白質・核酸の構造が収録されており、機能の異なる独立の蛋白質は 400 程度である。これらの蛋白質の立体構造から、配列の相同性もなく進化的にも機能的にも全く関係がないように見える蛋白質が同じ構造モチーフを有する例が多数明らかにされている。

#### 発明の開示

本発明の第一の課題は、あるアミノ酸配列を有する蛋白質のアミノ酸配列情報から、その蛋白質がとる可能性の高いスキヤッフォールドを推定することにより、立体構造をモデリングするための方法を提供することにある。正しいスキヤッフォールドの推定は、立体構造を正しく必要な精度でモデリングするための出発点となり得る。すなわち、本発明の究極の課題は、アミノ酸配列の情報のみから蛋白質の立体構造を高い信頼性をもって推定する方法を提供することにあり、その手段としてスキヤッフォールドを推定する方法を提供することが本発明の具体的課題である。また、本発明の別の課題は、上記の方法に有用なデータベースを提供することにある。

本発明者らは上記の課題を解決すべく鋭意努力した結果、立体構造が既知または推定可能な蛋白質について各アミノ酸残基の側鎖の環境情報を備えたデータベースを作成し、そのデータベースを利用することによって、立体構造未知の蛋白質のアミノ酸配列の情報から、その蛋白質がとる可能性の高いスキヤッフォールドを信頼性高く効率的に推定することができる方法を見出した。

すなわち本発明は、立体構造が既知又は推定可能な参考蛋白質のアミノ酸配列

に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキヤッフォールドを推定する方法を提供するものである。スキヤッフォールドを推定した後に、鋳型蛋白質と質問配列の最適マッチングに基づいて質問配列に対応した立体構造（3次元座標）を得ることができる。

本発明の好ましい態様によれば、参照蛋白質のアミノ酸配列が該参照蛋白質の立体構造に基づいて連続した2以上のアミノ酸残基からなる2以上の部分配列に分割された上記方法；参照蛋白質のアミノ酸配列が疎水コアの形成に実質的に関与する1又は2以上のコア部分配列と疎水コアの形成に実質的に関与しない1又は2以上のサブ部分配列とに分割された上記方法；参照蛋白質における各アミノ酸残基の側鎖の蛋白質内部への埋没度及び／又は蛋白質表面への露出度の情報と、質問配列の各アミノ酸の疎水性及び／又は親水性の性質とに基づいてマッチングを行う上記方法；参照蛋白質の1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の片端又は両端においてはギャップを考慮せずにマッチングを行う上記方法；並びに、ギャップが1又は2以上のアミノ酸残基の削除又は付加である上記方法が提供される。

本発明のさらに好ましい態様によれば、マッチングが以下の工程：

- (a) 1又は2以上のコア部分配列を質問配列上でスライドさせながら、必要に応じて該コア部分配列の片端又は両端においてはギャップを考慮してマッチングを行う工程（ただし、2以上のコア部分配列を用いる場合には、該コア部分配列は参照蛋白質のアミノ酸配列での出現順に順番に配置する）；及び
- (b) 工程(a)でコア部分配列にマッチングされなかった質問配列の部分につき、該参照配列における各コア部分配列とのつながりにしたがって、必要に応じてギャップを考慮しつつサブ部分配列を質問配列上でスライドさせる工程を含む上記方法；参照蛋白質のアミノ酸残基の側鎖に関する環境情報と質問配列上の対応ア

ミノ酸残基の疎水度パラメータとから計算したスコアに基づいて最適マッチングを選択する上記方法；並びに、参照蛋白質についての自己マッチングスコアを用いて上記スコアを規格化する工程を含む上記方法が提供される。

別の観点からは、本発明により、立体構造が既知又は推定可能な1又は2以上の参照蛋白質のアミノ酸残基の側鎖についての環境情報を含むデータベースであって上記の各方法に用いられるためのデータベースが提供される。このデータベースは、通常の記憶媒体、例えば磁気ディスク、光ディスク、CD-ROM、磁気テープなどに格納された形態で流通可能であり、該環境情報と質問配列のアミノ酸残基の性質との一致度をスコアとするマッチングによって、質問配列から構成される蛋白質の安定なスキヤッフォールドを推定するために用いることができる。

#### 図面の簡単な説明

第1図は、「立体構造」、「構造モチーフ」、及び「スキヤッフォールド」の関係を示した図である。図中、(a) は立体構造を示し、(b) は構造モチーフを示し、(c) はスキヤッフォールドを示す。

#### 発明を実施するための最良の形態

本発明の方法は、質問配列から理論的に可能な主鎖の折り畳み方を網羅・探索して評価するかわりに、参照蛋白質のデータベースから質問配列の蛋白質が立体構造的に類似する鋳型蛋白質を選び、鋳型蛋白質のスキヤッフォールドから質問配列の蛋白質の立体構造を推定することを特徴としている。本発明の方法は、一般的には、ワークステーション、パーソナルコンピューターなどの汎用のコンピューターを用いて高速に行うことができる。

本明細書において用いられる用語は、以下に述べる概念を含めて、最も広義に解釈する必要がある。「立体構造」とは3次元座標で表される蛋白質の構造を意味しており、アミノ酸残基の側鎖など存在する全原子を含む場合もあるが、それらの一部を省略することもある。「モデリング」とは、立体構造が実験的に解明

されているか否かに係わらず、ある蛋白質について存在可能性の高い立体構造を構築して原子の3次元座標で表現することをいう。

「2次構造」、「構造モチーフ」、及び「スキヤッフォールド」などの用語は Branden 及び Tooze らの著書に記載されている (Carl Branden and John Tooze, *Introduction to Protein Structure*, Garland publishing Inc. New York, 1991: 和訳「タンパク質の構造入門」、教育社、1992)。「構造モチーフ」と「スキヤッフォールド」は、ともにペプチド主鎖のみのトポロジーを表す点では共通しているが、「構造モチーフ」が2次構造群の組み合わせとペプチド鎖の流れを平面的かつ模式的に表現するのに対して、「スキヤッフォールド」は2次構造群の集合状態を含む蛋白質の3次元構造の骨組を意味する。「立体構造」、「構造モチーフ」、及び「スキヤッフォールド」の関係を図1に示す。

立体構造を推定したいアミノ酸配列を「質問配列」と呼び、その質問配列から構成される蛋白質を「質問配列の蛋白質」と呼ぶ。立体構造が既知または推定可能な蛋白質であって本発明のデータベースに含まれる蛋白質を「参照蛋白質」と呼び、参照蛋白質のうち、質問配列とのマッチングスコアがよく、質問配列の蛋白質と立体構造の類似性が高い蛋白質として選別された参照蛋白質を「鋳型蛋白質」という（「鋳型候補蛋白質」と呼ぶ場合もある）。鋳型蛋白質のスキヤッフォールドは、質問配列の蛋白質の立体構造を構築する際の鋳型として用いられる。

一般に2以上の配列をアミノ酸残基の一致度や相同性が高くなるように対応づけて並置する操作又は結果は「アラインメント」と呼ばれるが（「一致度」とは配列間で対応する残基間の厳密な一致を意味しており、「相同性」とは同等または類似など曖昧さを含めた一致の割合を意味する）、本発明の方法において「マッチング」（又は「対応付け」）という場合には、アミノ酸残基の一致又は相同性によらず、環境情報と性質の一致で残基を対応づけて並置する操作又は結果を意味している。ある対応付けにおける参照蛋白質の各アミノ酸残基の環境と質問配列のアミノ酸残基の性質との一致度を「マッチングスコア」（又は単に「スコア」とよぶ。「環境情報」は、主として、参照蛋白質の立体構造における各アミノ酸

残基の側鎖の蛋白質表面への露出度や存在環境を示す。アミノ酸配列について用いられる「ギャップ」という用語は、2以上のアミノ酸配列間の対応付けにおいていずれか一方の配列に対応するアミノ酸残基がない部分を指し、一方の配列から見ると1又は2個以上のアミノ酸残基の挿入及び／又は欠失していることを意味する。

本発明の好ましい態様では、このデータベースに参照蛋白質毎に立体構造を反映して分割した2以上の部分配列の情報とアミノ酸残基毎の環境情報とを格納しておく。参照蛋白質の環境情報と質問配列中の対応アミノ酸残基の疎水度パラメータとから計算されるマッチングスコアに基づいて2つの配列をマッチングする。疎水度パラメータは20種のアミノ酸残基に対して予め数値化しておくことが望ましい。

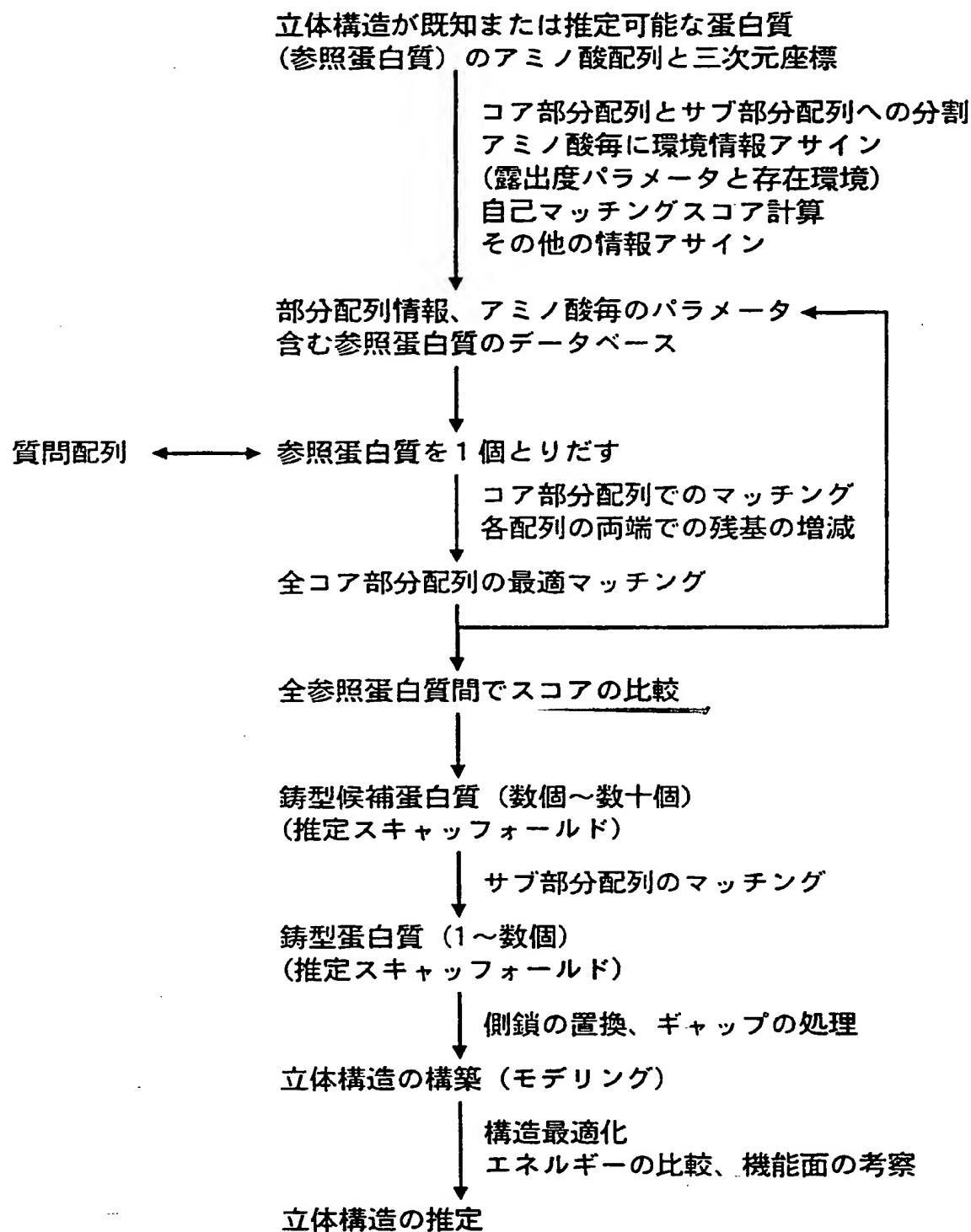
部分配列のうち、疎水コアの形成に関与するコア部分配列についてはギャップを入れずに質問配列上をスライドさせ、両端においてのみアミノ酸残基の増減(隣接するサブ部分配列の末端から1又は2個以上のアミノ酸残基をとって該コア部分配列に組み込むか、又はその逆の操作を意味する)を考慮してマッチングを行い、スコアのよい鋳型候補蛋白質を選別する。次に、疎水コアの安定化に関与しないサブ部分配列について必要に応じてギャップを考慮したマッチングを行い、鋳型候補の数を絞る。最終的な鋳型蛋白質の選択は、各蛋白質の最適マッチングのスコアを自己マッチングスコアで規格化して比べることにより行うことができる。

本発明の方法は、マッチングスコアが高いほど2つの配列間でスキヤッフオーラドや立体構造の類似性が高いとの仮定に基づいており、配列の相同性が殆どない配列間の残基のマッチングを蛋白質の折り畳み原理に従って行うことによって、参照蛋白質からの鋳型蛋白質の適切な選択を可能にしている点に特徴がある。ある。本発明の好ましい態様では、(1) データベースの作成；(2) 部分配列を用いたマッチング；(3) マッチングスコアの計算；(4) 参照蛋白質から鋳型候補蛋白質の選択；(5) 鋳型蛋白質の選択を主要な要素としている。

一般的な水溶性の単一サブユニットの蛋白質を構成するペプチド鎖については、アミノ酸残基の疎水性側鎖ができるだけ分子内部に埋もれて露出せず、一方、親水性側鎖はできるだけ蛋白質分子表面に露出するのが自由エネルギー的に有利である。もっとも、細胞膜に相互作用する蛋白質や複数のサブユニットからなる蛋白質など存在環境が異なる蛋白質では、個々の蛋白質構造やサブユニット構造だけをみると、疎水性側鎖が蛋白質表面に露出していることもある。本発明の方法では、このような蛋白質の立体構造の多様性も考慮されており、個々のアミノ酸残基の存在環境を反映したスコア付けができるようになっている。アミノ酸残基毎の特定2次構造のとり易さの概念（例えば P. Y. Chou, & G. D. Fasman, *Adv. enzymol.* 47, 45, 1978）は基本的には用いないが、必要に応じてそれらの概念を加味したスコアを用いることができる。

### 実施例

以下、本発明の方法の好ましい態様をスキームで示し、このスキームに従って本発明の方法をより具体的に説明するが、本発明の方法はこのスキームの細部又は以下の説明の細部に限定されることはない。



### (A) データベースの作成

立体構造が既知または推定可能な参照蛋白質について、立体構造に関する情報として、各アミノ酸残基の環境情報及び部分配列の情報を含むデータベースを作成しておく。データベースに収納する蛋白質としては、立体構造の情報が既知または推定可能な蛋白質であれば、すべてエントリーとすることができます。同時に構造決定された蛋白質中に複数のサブユニットが含まれるときは、独立のエントリーとしておくことが好ましい。ペプチド鎖で繋がった複数ドメインからなる構造については、全体構造とともに各ドメインも独立エントリーとしておくことが好ましい。

参照蛋白質毎に含める情報は以下の通りである。

#### (1) 一般的な情報

- (a) 蛋白質名（蛋白質コード）、サブタイプ、アミノ酸数、アミノ酸配列、ドメイン、サブユニットなどに関する情報。
- (b) 立体構造に関する情報として、立体構造の決定（又は推定）方法（結晶解析法、n m r 法、又はモデリング法のいずれにより立体構造を決定したか）、P D B コード、モデリング法による場合には鋳型として利用した蛋白質分子名、結晶解析の場合には共結晶化された分子がある場合にはその化学名などの情報のうち適宜のものを含めておく。
- (c) 生物学的意義に関する情報として、生物学的機能、生物種、存在する組織・臓器、エフェクター分子などの情報を含めておく。

#### (2) 部分配列情報

蛋白質の立体構造における特徴に従って、配列を分割して2以上の部分配列とし、各部分配列について以下の情報をデータベースに含めるのが望ましい。

- (a) N末端から何番目の部分配列か
- (b) コア部分配列かサブ部分配列かのフラグ
- (c) 始点と終点（N末端からのアミノ酸配列番号で）、配列の長さ、その距離及

びベクトル、サブ部分配列ならば最短の残基数

(d) 部分配列間の距離及びベクトル

(e)  $\beta$ -シート形成の水素結合をする相手の部分配列番号、および逆平行または平行  $\beta$ -シートなどの区別、あるいは一定の距離内にある部分配列番号など

(f) 自己マッチングスコア（後述）

部分配列への分割の基準は特に限定されないが、 $\alpha$ -ヘリックス構造や $\beta$ -ストランド構造などの2次構造を形成し実質的に疎水コアの形成に関わる部分をコア部分配列とし、それ以外の部分をサブ部分配列とすることができる。各部分配列は、例えば、7残基以上を含む連続した配列とすることが望ましい。 $\beta$ -ターン構造については、はじめからコアまたはサブ部分配列に加えても構わないし、別のフラグによって区別し、検索時にコアかサブかの判断を加えてもよい。分割の作業は、コンピュータグラフィックス画面上で対話的に個々の蛋白質について行ってもよいし、分割の基準を定めたプログラムを作成して自動的に行ってもよい。その判断基準として、例えば、隣り合う4つのCアルファ原子のなす擬ねじれ角などの数値を基準に用いれば、自動的に部分配列に分割できる。

### (3) アミノ酸残基毎の環境情報

#### (a) 露出度パラメータ

参照蛋白質の全てについて、立体構造に基づいて残基毎の蛋白質表面への側鎖の露出度及び蛋白質内部への側鎖の埋没度を計算し、その計算値に基づいて露出度パラメータを割り振る。本明細書において用いられる「露出度パラメータ」という用語は、各残基の側鎖がどの程度分子表面に露出しているか、又は埋没しているかを示す数値を意味している。露出度パラメータを定義する方法は特に限定されず、いかなる方法を採用してもよいが、露出度が高いものはマイナスの値、露出度が低いものはプラスの値をとるように設定するのが望ましい。例えば、立体構造において各アミノ酸側鎖の溶媒への接触可能表面と蛋白質原子への接触表面積を計算して、その差から露出度パラメータを算出して利用することができる。

また、例えば、全分子表面に対する溶媒接触表面の割合を基準として定めることもできる。

(b) 存在環境フラグの設定

細胞膜に相互作用する蛋白質では膜と相互作用する部分の表面に疎水性アミノ酸残基の側鎖が露出しており、また、サブユニットやドメイン構造が集合して安定化する接触面を有する蛋白質では、接触面に疎水性アミノ酸残基の側鎖が露出している場合があり、これらの蛋白質は単独で存在する水溶性蛋白質とは異なる性質を有している。このような蛋白質については、一般的には、先述の折り畳み原理はそのままでは適用できない。データベース作成時に、例えば、次のような存在環境フラグを露出度パラメータとは別に各アミノ酸残基に与えておくことができる。

例えば、蛋白質の由来や機能に関する実験結果および立体構造から、各アミノ酸の存在環境が以下のいずれに相当するかを推定して環境フラグを与えておき、マッチングやスコアの計算の際に考慮することが可能である。

0 : 不明 (未定義または定義できず)

1 : 分子内部 (蛋白質内、サブユニット内、ドメイン内の接触)

2 : 分子内孔 (リガンド結合部位)

3 : 分子表面 (水環境と接触)

4 : 分子表面 (別蛋白質、別サブユニット、別ドメインと接触)

5 : 分子表面 (膜と接触)

さらに、立体構造形成に特殊な影響を与えるアミノ酸残基については、特殊残基であることを示すフラグを与えておき、マッチングやスコアの計算の際に考慮することができる。例えば、S-S結合しているシステイン残基やプロリンのように主鎖に水素結合性官能基が不足しているアミノ酸残基、または側鎖原子間に強い親水性相互作用を形成する可能性のある残基などに適用することができる。

### (B) 質問配列と参照蛋白質の配列間のマッチング

一方の配列を他方の配列上スライドさせて最適のマッチングを効率よく探すために、部分配列の概念を利用することができる。そのために、上記データベースは、各参照蛋白質について立体構造から得られる部分配列の情報とアミノ酸残基ごとの環境情報がアミノ酸配列順に収めてある。一方、質問配列が有している情報はアミノ酸配列の情報だけであり、アミノ酸残基毎の疎水度パラメータの表からとった値を当てはめてスコアの計算に用いる。データベースから参照蛋白質を1つずつ取り出し、アミノ酸配列中に出現する順に並べた部分配列群を質問配列上でスライドさせて、部分配列群と質問配列との間でマッチングスコアの最もよいマッチングを探す。

#### (1) 部分配列を用いたマッチング

部分配列を用い、かつコア部分配列ではギャップを考慮しないでアミノ酸残基単位でマッチングを行うことにより、最適マッチングの配列間対応付けを高速に検索することができる。一般に、進化の過程では残基の置換とともに挿入や欠失が起きることが多く、その考慮は配列のマッチングに不可欠である（「従来の技術」の欄を参照）。しかしながら、一般的には、ギャップが入るのはサブ部分配列であることが多い。これは、疎水コアの安定化に関与するコア部分配列でその両端以外で挿入や欠失が起きると、その安定スキヤッフォールド自体が損なわれ、蛋白質の立体構造が大きく変化してしまうからである。

そこで、本発明の方法の好ましい態様では、コア部分配列とサブ部分配列とに分けて2段階のマッチングを行う。第一段階のコア部分配列を用いたマッチングにおいては、まずギャップを考慮せずに1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の両端においてのみアミノ酸残基の増減を考慮したマッチングを行って最適マッチングを探索する。

各コア部分配列について質問配列の上をスライドさせながらマッチングスコアを計算して保存する。全部のコア部分配列について同様に計算した後、全体とし

て最適なマッチングを決定する。2以上のコア部分配列を用いる場合には、質問配列上に2以上のコア部分配列を参照蛋白質のアミノ酸配列中での出現順序に従って重なり合わないように配置し、コア部分配列間には4個程度以上のアミノ酸残基の存在を仮定して（立体的に結合可能なアミノ酸残基数で隣のコア部分配列と順番に繋がるという条件、例えば $\beta$ -ターンなどに要するアミノ酸残基数）、それぞれのコア部分配列の順序を変えずに質問配列上をスライドさせ、最も総スコアのよいマッチングを選択する。この際、各コア部分配列のマッチングスコアが最大である必要はない。この第一段階で、高いスコアを与える鋳型蛋白質を数個から数十個選択し、スキャッフオールド候補として第二段階に進む。

### (C) マッチングスコア

マッチングスコアは参照蛋白質の環境情報の露出度パラメータ  $EP(i)$  と質問配列の残基の疎水度パラメータ  $HB(j)$  を用いて計算する。 $i$  は参照蛋白質のアミノ酸配列中のアミノ酸残基番号であり、 $j$  はそれと対応づけられた質問配列の残基番号を指す。マッチングスコアの計算式は、参照蛋白質の分子内部に埋没した側鎖環境に質問配列の疎水性の強いアミノ酸残基が対応し、分子表面に露出した側鎖環境に質問配列の親水性の強いアミノ酸残基が対応すると高いスコアが得られるような計算式であれば、いかなるものを利用してもよい。マッチングスコアは例えば、最も単純には次式によって計算することができる。

$$\text{残基毎のマッチングスコア} = EP(i) \times HB(j)$$

部分配列のマッチングスコア = 配列に含まれる残基のマッチングスコアの和

全配列のマッチングスコア = 全部分配列のマッチングスコアの和

### (1) 疎水度パラメータ

20種のアミノ酸残基のそれぞれに疎水性又は親水性の性質に関連した疎水度パラメータを与えておく。疎水度パラメータの決定方法は特に限定されず、いかな

る基準による値を用いてもよい。例えば、文献に記載されたアミノ酸毎の疎水性値を用いてもよく、又は適宜の方法により独自の基準で算出したものを用いてもよい。また、あるアミノ酸について、結晶解析された蛋白質中の全出現回数に対して蛋白質分子内部に埋没された残基の比率を統計的に求めておき、その比率を該アミノ酸の疎水度パラメータとして利用してもよい。

また、例えば、個々のアミノ酸残基に別々の値を与えてよいが、次のように段階化した疎水度パラメータを与えることもできる。

表 1

- 2 : 強い疎水性 (イソロイシン、バリン、ロイシン、フェニルアラニン)
- 1 : 弱い疎水性 (アラニン、メチオニン、シスチン、チロシン)
- 0 : ほぼ中性 (グリシン、プロリン、リジン、アルギニン)
- 1 : 弱い親水性 (スレオニン、ヒスチジン)
- 2 : 強い親水性 (セリン、アスパラギン、アスパラギン酸、グルタミン、グルタミン酸)

## (2) 自己マッチングスコア

アミノ酸数及びアミノ酸組成の異なる蛋白質間で質問配列へのマッチングの良さを比較するためにはスコアの規格化をしておくことが望ましい。そのために、各参照蛋白質について、それ自体のアミノ酸配列の露出度パラメータ EP(i) とデータベースに用意された環境情報の疎水度パラメータ HB(i) とからマッチングスコアを予め計算してデータベースに保存しておく。例えば、次式のように計算するとよい。

$$\text{自己マッチングスコア} = \sum (EP(i) \times HB(i))$$

全部分配列の質問配列へのマッチングが終了して最適マッチングが得られた後、

得られたマッチングスコアに自己マッチングスコアを乗じて規格化することができる。全参照蛋白質について規格化された最適マッチングスコア同士を比較して最適な鋳型候補蛋白質を選択することができる。自己マッチングスコア及びマッチングスコアは、アミノ酸残基数が多いほど大きな値を取りやすい。

#### (D) 鋳型候補蛋白質の選択

鋳型候補蛋白質の選択の手順は概ね以下の通りである。

- (a) データベースから参照蛋白質を1個づつ取り出し、質問配列に対しマッチングを行う；
- (b) コア部分配列につきギャップを考慮せずに質問配列上をスライドしマッチングスコアを算出する；
- (c) 必要に応じて、各コア部分配列のN末端またはC末端の残基を増減しながらマッチングし、最適マッチングを得る；
- (d) 参照蛋白質全部について工程(a)～(c)を行い、最適マッチングとマッチングスコアを得る；
- (e) 規格化したスコアにより参照蛋白質からスコアのよいものを鋳型候補蛋白質とする（この段階で、それらの構造は質問配列のスキヤッフォールドの候補とすることができる）；
- (f) コア部分配列間をつなぐサブ部分配列のマッチングを行う。質問配列の対応する配列部分との配列長の違いやギャップの存在を考慮し、最適のマッチングとマッチングスコアを得る；及び
- (g) 規格化したスコアによって鋳型蛋白質を選択する。

#### (E) 立体構造の構築

質問配列の蛋白質の立体構造は、鋳型蛋白質のスキヤッフォールドの構造、及び該参照蛋白質と質問配列との最適マッチングの結果に基づいて、アミノ酸残基の側鎖の置換を行うことにより構築することができ、質問配列に対応した3次元座

標を得ることができる。優劣をつけ難い2以上の鋳型蛋白質がある場合には、すべてについて立体構造を構築することが望ましい。サブ部分配列の長さが鋳型候補と異なる場合には、結晶構造に現れたループ構造を集めたデータベース等を用いて、該サブ部分配列の適切なトポロジーを決定することができる。鋳型のスキヤツフォールドが修正される部分については、マッチングスコアと同様に残基の性質と露出度を考慮しつつトポロジーを決定することができる。重大な立体障害や立体構造を不安定化する分子内部の隙間などの有無を確認し、また構造最適化計算や分子動力学計算により構造の微調整を行ったのちに立体構造の安定性を比較する。

最後に、全ての鋳型に基づいて構築された立体構造をエネルギーとマッチングスコアにより順位付けするが、該質問配列の蛋白質の機能が既知の場合には、該鋳型蛋白質に知られた機能との対応、その機能の発現に関与するものと推定されるアミノ酸残基の立体構造上の位置の妥当性、さらにはアミノ酸変異による機能への影響などの情報を鋳型蛋白質の選別に利用できる。

### 産業上の利用可能性

本発明の方法によれば、立体構造が既知または推定可能な蛋白質のアミノ酸配列データベースに基づいて、アミノ酸配列の情報のみからそのアミノ酸配列により構築される蛋白質の立体構造に関する情報を信頼性高く効率的に入手することができる。

## 請求の範囲

1. 立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキップオールドを推定する方法。
2. 参照蛋白質のアミノ酸配列が該参照蛋白質の立体構造に基づいて連続した2以上のアミノ酸残基からなる2以上の部分配列に分割された請求の範囲第1項に記載の方法。
3. 参照蛋白質のアミノ酸配列が疎水コアの形成に実質的に関与する1又は2以上のコア部分配列と疎水コアの形成に実質的に関与しない1又は2以上のサブ部部分配列とに分割された請求の範囲第2項に記載の方法。
4. 参照蛋白質における各アミノ酸残基の側鎖の蛋白質内部への埋没度及び／又は蛋白質表面への露出度の情報と、質問配列の各アミノ酸の疎水性及び／又は親水性の性質とに基づいてマッチングを行う請求の範囲第1項ないし第3項のいずれか1項に記載の方法。
5. 参照蛋白質の1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の片端又は両端以外においてはギャップを考慮せずにマッチングを行う請求の範囲第1項ないし第4項のいずれか1項に記載の方法。
6. ギャップが1又は2以上のアミノ酸残基の削除又は付加である請求の範囲第5項に記載の方法。
7. マッチングが以下の工程：
  - (a) 1又は2以上のコア部分配列を質問配列上でスライドさせながら、必要に応じて該コア部分配列の片端又は両端においてはギャップを考慮してマッチングを行う工程（ただし、2以上のコア部分配列を用いる場合には、該コア部分配列は

参照蛋白質のアミノ酸配列での出現順に順番に配置する) ; 及び

(b) 工程(a)に続いて、必要に応じてギャップを考慮しつつサブ部分配列を質問配列上でスライドさせる工程を含む請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

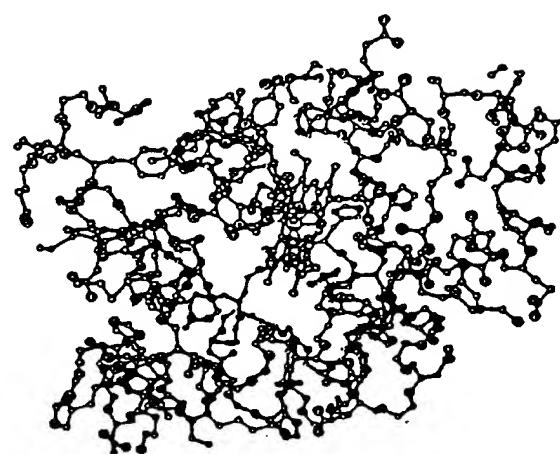
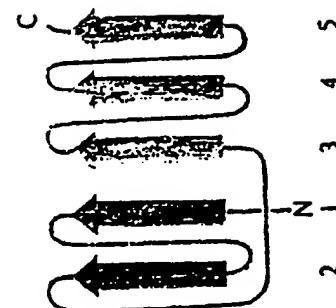
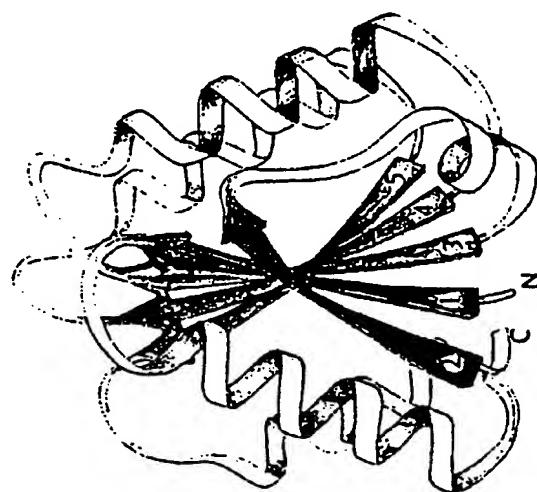
8. 参照蛋白質のアミノ酸残基の側鎖に関する環境情報と質問配列上の対応アミノ酸残基の疎水度パラメータとから計算したスコアに基づいて最適マッチングを選択する、請求の範囲第1項ないし第7項のいずれか1項に記載の方法。

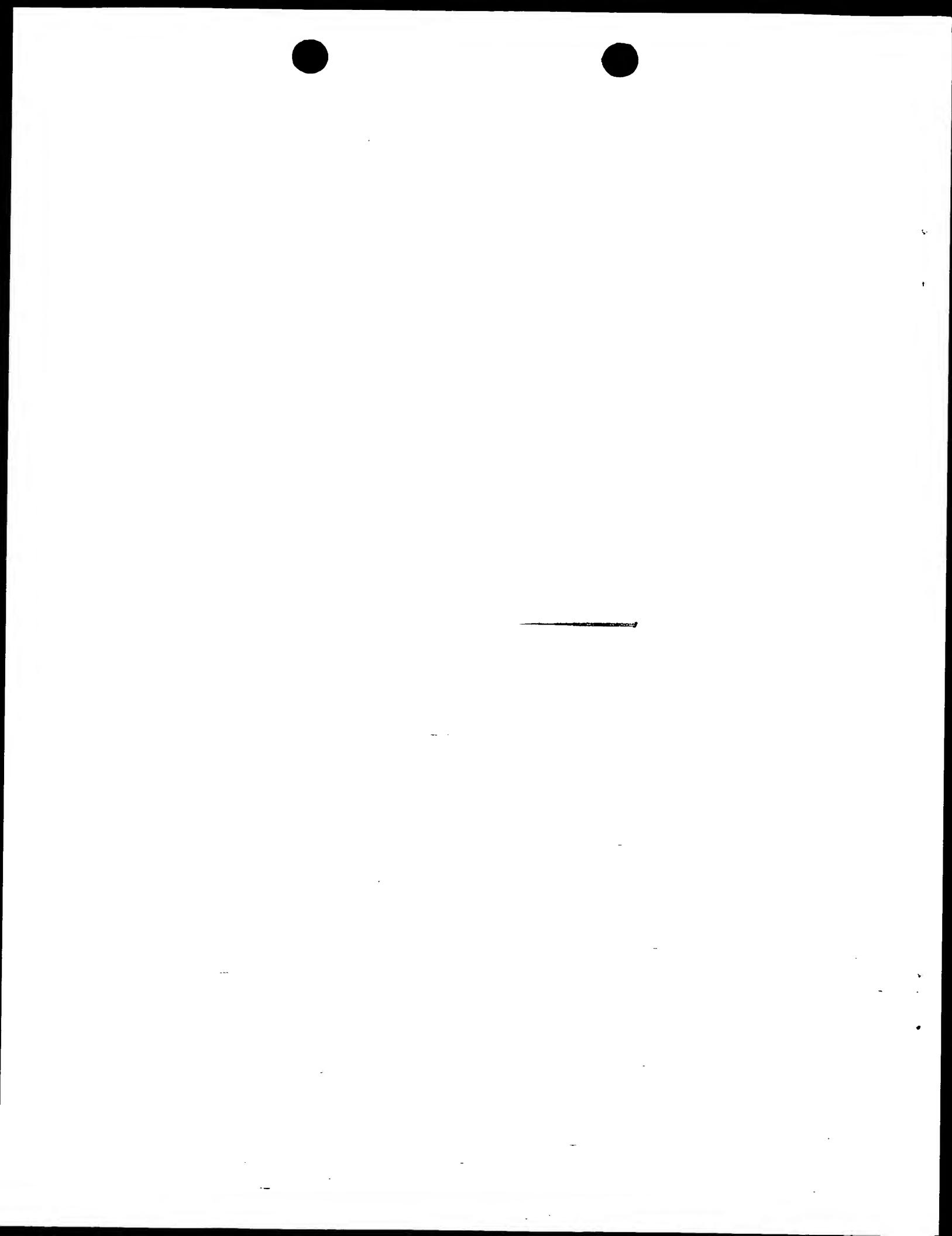
9. 参照蛋白質についての自己マッチングスコアを用いて上記スコアを規格化する工程を含む請求の範囲第8項に記載の方法。

10. さらに質問配列の蛋白質の立体構造を構築する工程を含む請求の範囲第1項ないし第9項のいずれか1項に記載の方法。

11. 立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースであって、請求の範囲第1項ないし第10項のいずれか1項の方法に用いるためのデータベース。

第1図





## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/04457

A. CLASSIFICATION OF SUBJECT MATTER  
Int.Cl<sup>6</sup> G01N33/68, G06F17/00, G06F15/18, C07K14/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
Int.Cl<sup>6</sup> G01N33/68, G06F17/00, G06F15/18, C07K14/00Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-1998  
Kokai Jitsuyo Shinan Koho 1971-1998 Jitsuyo Shinan Toroku Koho 1996-1998

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PA	JP, 10-95796, A (NEC Corp.), 14 April, 1998 (14. 04. 98) (Family: none)	1-10
X	JP, 7-206894, A (Advanced Technology Institute K.K.), 8 August, 1995 (08. 08. 95) (Family: none)	1-10
A	JP, 8-69446, A (NEC Corp.), 12 March, 1996 (12. 03. 96) (Family: none)	1-10
A	JP, 7-105179, A (NEC Corp.), 21 April, 1995 (21. 04. 95) (Family: none)	1-10
A	JP, 7-93286, A (NEC Corp.), 7 April, 1995 (07. 04. 95) (Family: none)	1-10
A	JP, 7-13959, A (NEC Corp.), 17 January, 1995 (17. 01. 95) (Family: none)	1-10

 Further documents are listed in the continuation of Box C.  See patent family annex.

• Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search  
14 December, 1998 (14. 12. 98)Date of mailing of the international search report  
6 January, 1999 (06. 01. 99)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/04457

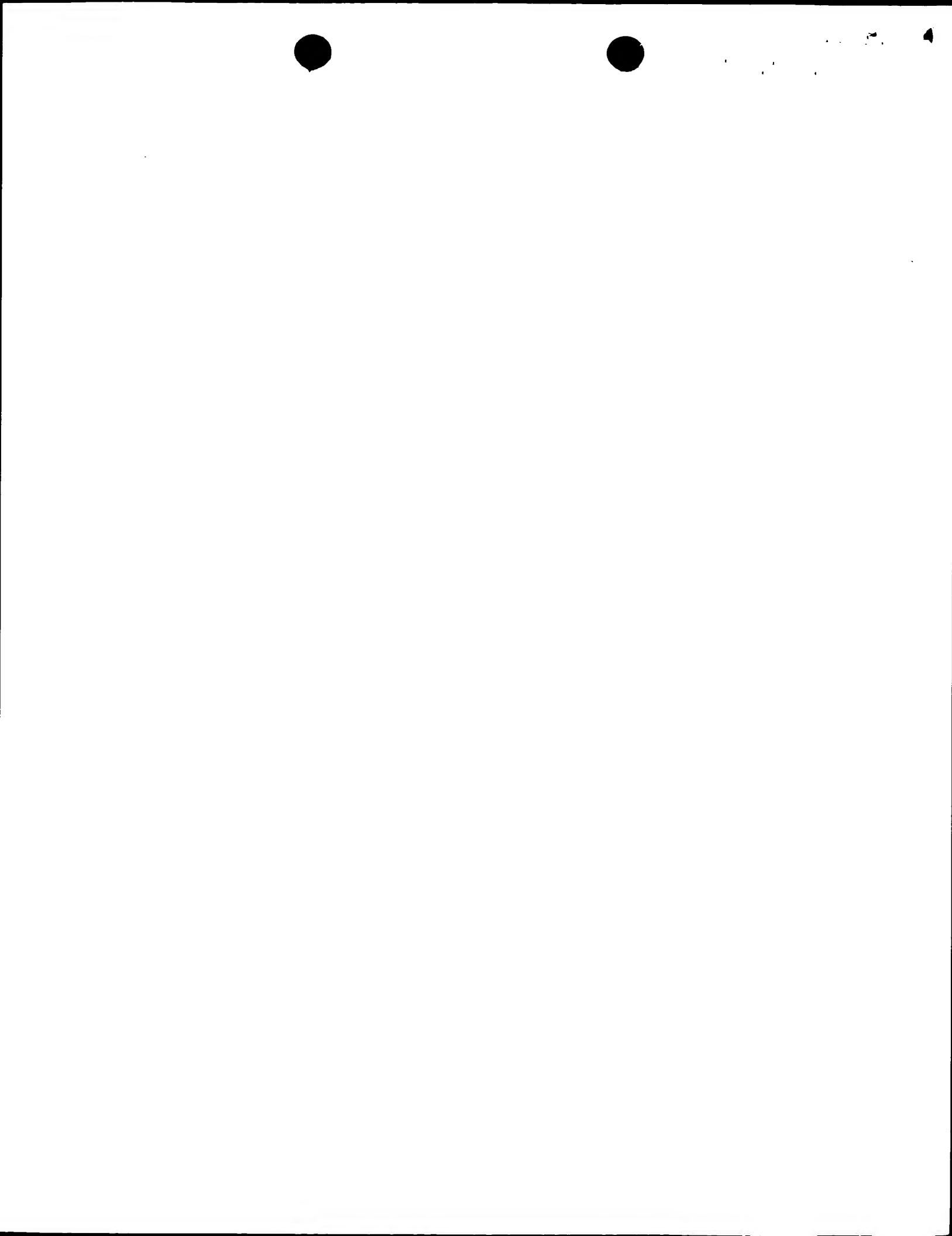
## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP, 5-282383, A (Fujitsu Ltd.), 29 October, 1993 (29. 10. 93) (Family: none)	1-10
A	JP, 4-45781, A (Fujitsu Ltd.), 14 February, 1992 (14. 02. 92) (Family: none)	1-10

立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列を疎水コアの形成に実質的に関与するコア部分配列と関与しないサブ部分配列に分割し、各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、部分配列毎に参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキヤッフォールドを推定する方法。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE アラブ首長国連邦	ES スペイン	LI リヒテンシュタイン	SG シンガポール
AL アルバニア	FI フィンランド	LK スリ・ランカ	SI スロヴェニア
AM アルメニア	FR フランス	LR リベリア	SK スロヴァキア
AT オーストリア	GA ガボン	LS レソト	SL シエラ・レオネ
AU オーストラリア	GB 英国	LT リトアニア	SN セネガル
AZ アゼルバイジャン	GD グレナダ	LU ルクセンブルグ	SZ スウェーデン
BA ボスニア・ヘルツェゴビナ	GE グルジア	LV ラトヴィア	TD チャード
BB バルバドス	GH ガーナ	MC モナコ	TG トーゴ
BE ベルギー	GM ガンビア	MD モルドバ	TJ タジキスタン
BF ブルギア・ファソ	GN ギニア	MG マダガスカル	TM トルクメニスタン
BG ブルガリア	GW ギニア・ビサオ	MK マケドニア旧ユーゴスラヴィア	TR トルコ
BJ ベナン	GR ギリシャ	共和国	TT トリニダッド・トバゴ
BR ブラジル	HR クロアチア	ML マリ	UA ウクライナ
BY ベラルーシ	HU ハンガリー	MN モンゴル	UG ウガンダ
CA カナダ	ID インドネシア	MR モーリタニア	US 米国
CF 中央アフリカ	IE アイルランド	MW マラウイ	UZ ウズベキスタン
CG コンゴー	IL イスラエル	MX メキシコ	VN ヴィエトナム
CH スイス	IN インド	NE ニジェール	YU ユーゴースラビア
CI コートジボアール	IS アイスランド	NL オランダ	ZA 南アフリカ共和国
CM カメルーン	IT イタリア	NO ノルウェー	ZW ジンバブエ
CN 中国	JP 日本	NZ ニュージーランド	
CU キューバ	KE ケニア	PL ポーランド	
CY キプロス	KG キルギスタン	PT ポルトガル	
CZ チェコ	KP 北朝鮮	RO ルーマニア	
DE ドイツ	KR 韓国	RU ロシア	
DK デンマーク	KZ カザフスタン	SD スーダン	
EE エストニア	LC セントルシア	SE スウェーデン	



## 明細書

## 蛋白質の立体構造の推定方法

## 技術分野

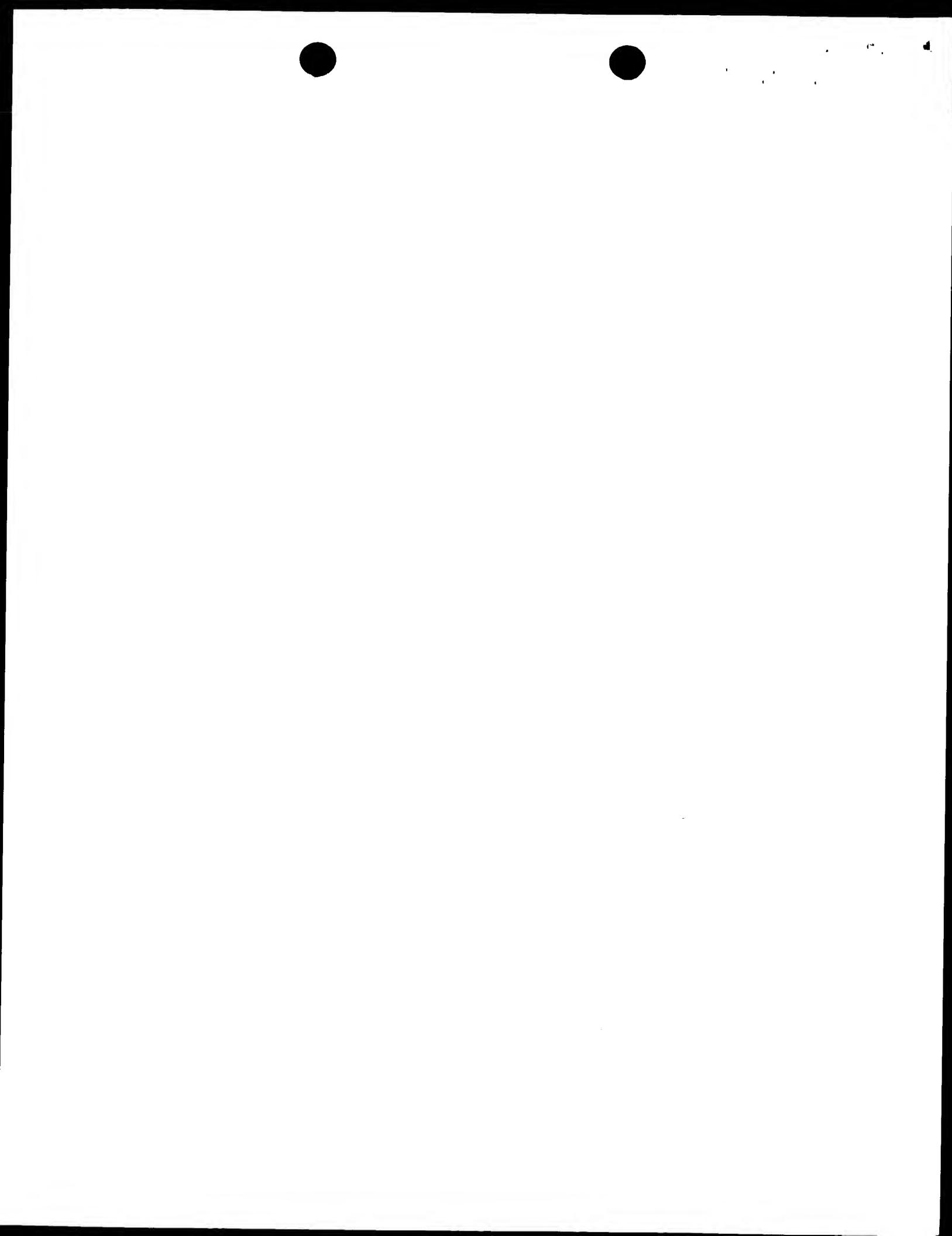
本発明は、蛋白質の立体構造の推定方法に関するものである。

## 背景技術

アミノ酸配列から蛋白質の立体構造を推定することは理論的に不可能とはいえない。しかしながら、現在のところ、配列情報から蛋白質の立体構造を確実に推定する手段は開発されておらず、蛋白質の立体構造を知るための手段は、X線結晶構造解析やnmr解析などの実験的方法に限られている。蛋白質の立体構造の情報は、その機能を原子レベルで理解し、またその蛋白質を標的とする医薬の創製やさらに優れた機能をもつ有用な蛋白質の創製に不可欠である。近年、遺伝子情報の解析手段が急速に進歩した結果、実際に蛋白質が単離されないまま配列情報だけが解明される例が急増している。従って、配列情報から蛋白質の立体構造や機能を推定する有効な手段の開発が切望されているのが現状である。

あるアミノ酸配列を有する蛋白質の存在がわかった場合、配列データベースから相同性のある蛋白質を検索するのが一般的である。アミノ酸配列の一致度がある程度よい蛋白質が見つかると、さらにその蛋白質との間で相同性やギャップも考慮したアラインメントを行い、さらに相同性の高いアラインメントの探索が行われる。目的の蛋白質と機能既知の蛋白質との相同性が高い場合には機能がその既知蛋白質に類似しており、一方、立体構造既知の蛋白質との相同性が高い場合には立体構造がその蛋白質に類似しているとの推定が成り立つ。また、相同性が高ければ高いほど機能や立体構造が類似している確率が高く、その推定の信頼性が高いと考えられる。

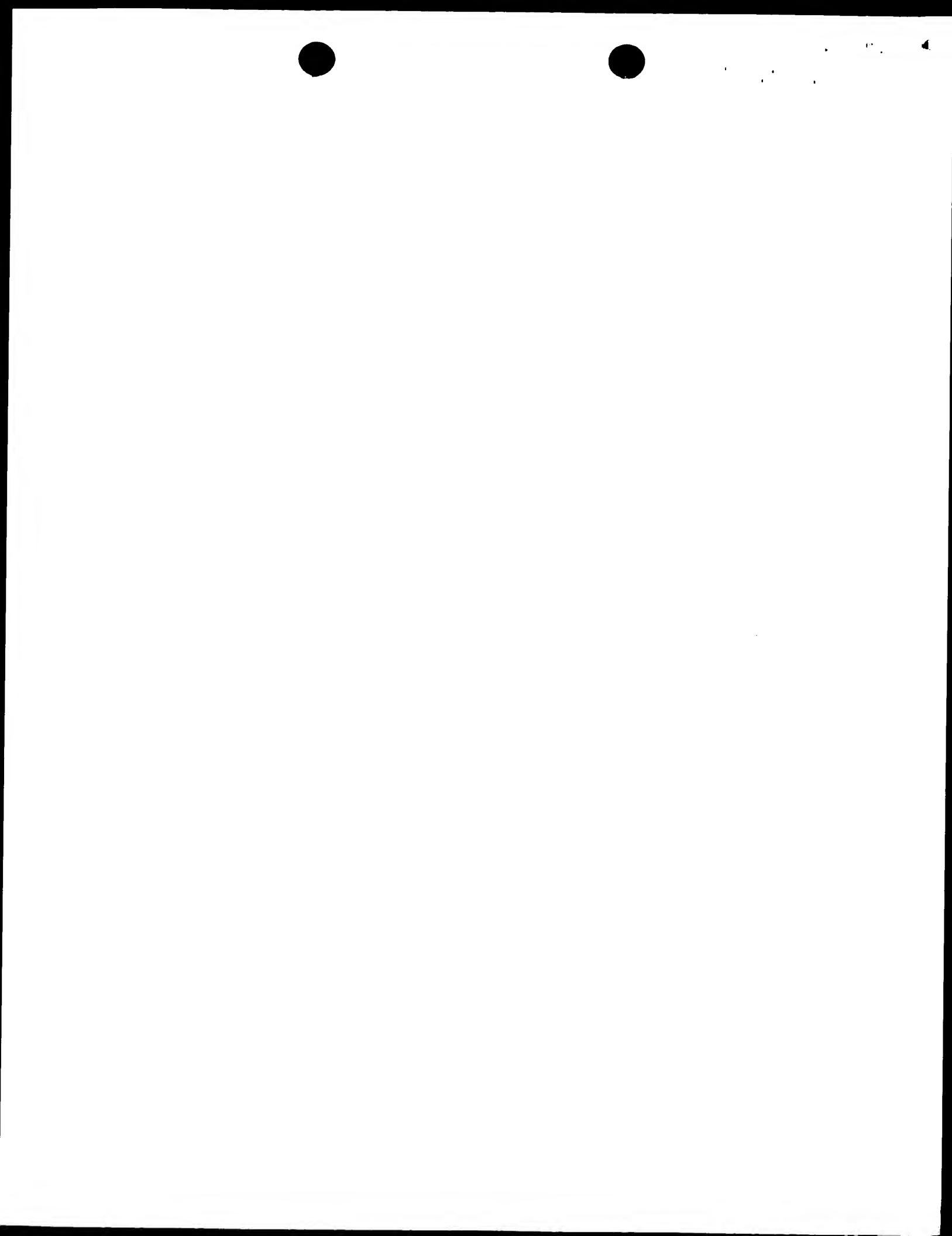
立体構造既知の蛋白質の配列とある程度（一般に30%程度）以上の相同性が



認められた場合には、その立体構造を鋳型として立体構造を構築するホモジーモデリング法が行われる。鋳型の立体構造に基づいて対応付けられた残基が鋳型と異なる場合には、側鎖を置換することによって仮想的な立体構造を構築することができる。アラインメント上のギャップは、鋳型立体構造中に対応するアミノ酸がないか、あるいは鋳型の方が余計なアミノ酸をもつことになるので別途処理する必要があるが、ギャップの存在は鋳型に基づくモデリング作業を難しくし、かつ信頼性を損なうので、できるだけギャップの数を少なくするようにギャップにペナルティを与えるアラインメント法が推奨されている。

問題のアミノ酸配列とある程度以上相同性の高いアミノ酸配列を有する立体構造既知の蛋白質が見つからない場合には、ホモジーモデリングは不可能である。一方で、蛋白質の結晶構造情報が蓄積されるにつれて、アミノ酸配列の相同性が殆どなく機能も全く異なる複数の蛋白質が、類似の立体構造を持つ例が多数わかつってきた。このことは、蛋白質が安定な立体構造をとるための物理化学的要因を考慮すれば、アミノ酸配列の相同性が低い場合にも、立体構造既知の蛋白質群中から鋳型となる立体構造を見出せる可能性があることを示している。

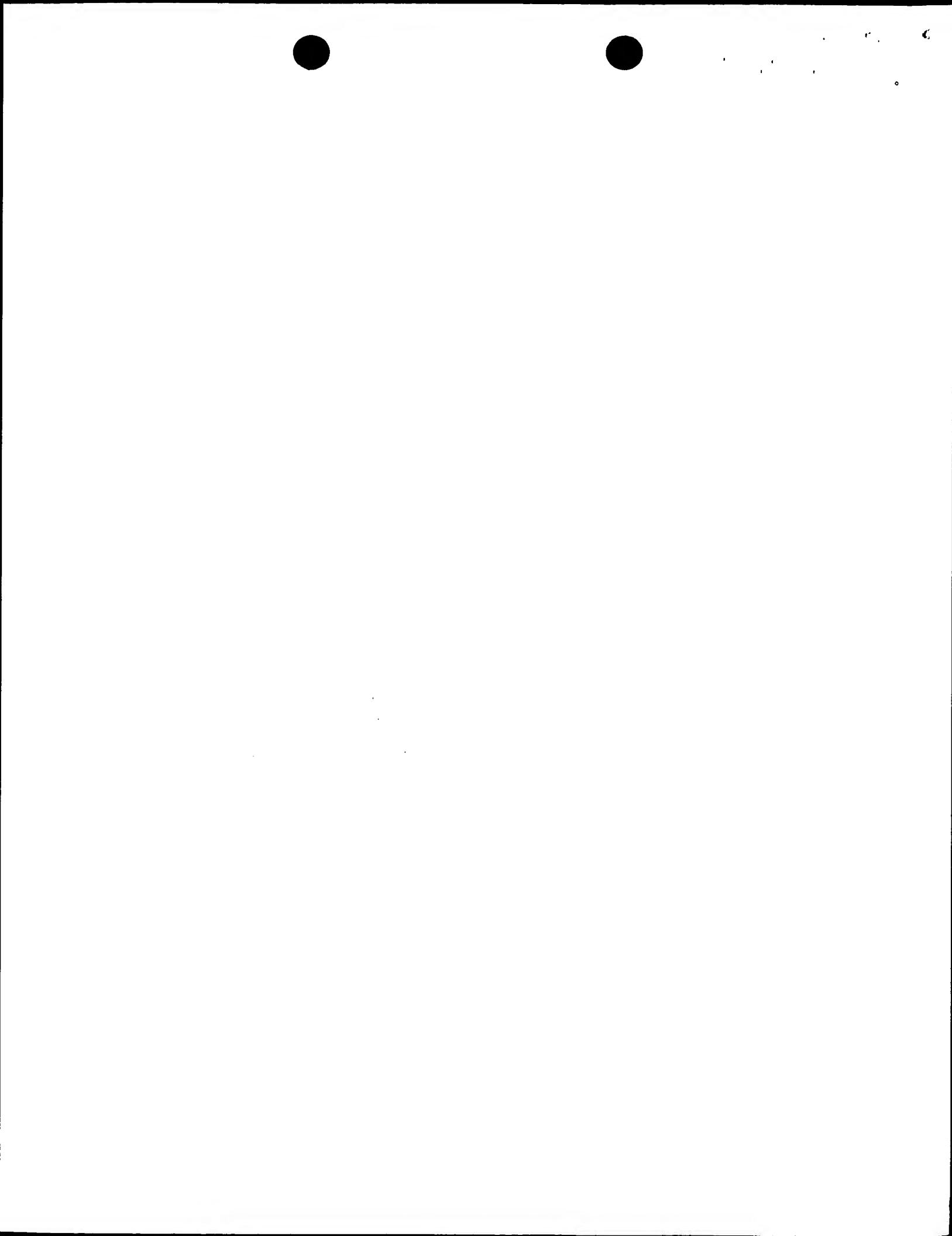
近年、アミノ酸残基毎の疎水性など物理的な性質の一致を考慮したスコアを用いることにより、アミノ酸配列の相同性が低くても立体構造的に類似性が高いと推定される鋳型蛋白質を立体構造既知の蛋白質群から選ぶ方法が開発された。代表的な方法として、Eisenberg らによる 3D-1D 法 (R. Luthy, J. U. Bowie and D. Eisenberg, *Nature*, 356, 83, 1992) がある。この方法は、アミノ酸配列の相同性に加え、立体構造既知の蛋白質について各アミノ酸残基の属する 2 次構造とその残基位置の環境を表すパラメータと、問題の配列の各アミノ酸残基に対して 2 次構造毎に与えたパラメータとを用いて、対応づけられたアミノ酸残基間で類似性のスコアを計算する工程を含んでいる。この方法では、蛋白質のペプチド鎖の折り畳み方についての膨大な自由度の問題を既知の結晶構造群を鋳型として用いることで回避しており、評価の要素として疎水性などの物理的パラメータを含めることによって配列の相同性が低い場合にもモデリングが可能である。



しかしながら、立体構造が類似している場合であっても、アミノ酸残基数、2次構造、及びループの長さが同じことは少ないので、アミノ酸配列間の対応づけに基づいた 3D-1D 法を実際に適用する場合には数多くの問題が生じる。例えば、アミノ酸配列間の単純なスライド（スレッディング）に加え、どちらかの配列に部分的な配列の欠損など（ギャップ）を考慮して対応付けする必要があるが、ホモロジーモデリングと同様にギャップの導入はモデリングの信頼性を低下させるてしまう。配列の相同性が低いときに、必要にして最小限のギャップを考慮した対応付けをいかにして実現するかが問題である。また、疎水性親水性などのパラメータの他、20種類のアミノ酸残基に対して2次構造毎のパラメータを用いるなど非常に多数のパラメータに依存しており、パラメータの改良による予測性の向上は期待しにくい。

アミノ酸配列から蛋白質の立体構造を予測する研究の歴史は、配列のどの部分がどのような2次構造をとるかを予測することから始まった。多数の蛋白質の結晶構造情報から統計的に求められたアミノ酸残基毎あるいは連続した数アミノ酸残基の組毎に、 $\alpha$ -ヘリックス又は $\beta$ -シートになり易さを示すパラメータを用いて、問題のアミノ酸配列から顕著な傾向を示す連続した領域を検出し、それがどちらの構造をとるかを予測するものである。その代表的なものとして、Chou と Fasman らによる 2 次構造予測法 (P. Y. Chou, & G. D. Fasman, *Adv. enzymol.* 47, 45, 1978) がある。しかしながら、このような方法は、2 次構造の 3 次元的集合に関してはなんら情報を与えるものではなく、また予測された 2 次構造と結晶解析で確認された 2 次構造の一致が 60% 前後であることから、立体構造の推定方法としてはほとんど利用価値がない。

蛋白質の安定な折り畳み構造を、純粹に計算的手法によって先入観を入れずに予測する試みも行われるようになった（いわゆる *ab initio* 予測法）。しかしながら、蛋白質は極めて自由度が大きい分子であり（100 残基程度の蛋白質でも考慮すべき自由度のパラメータは 400 以上である）、すべての自由度を考慮して可能な構造を十分に探索することは、現在利用可能なコンピュータでは不可能で



ある。また、可能な立体構造の安定性を正しく評価できるほど蛋白質の構造の安定化に関わるファクター（例えば、水の物理化学的性質、疎水相互作用、静電相互作用）についての研究は進んでいないなどの理由から、このような構造予測法の成功は今のところ期待できない。

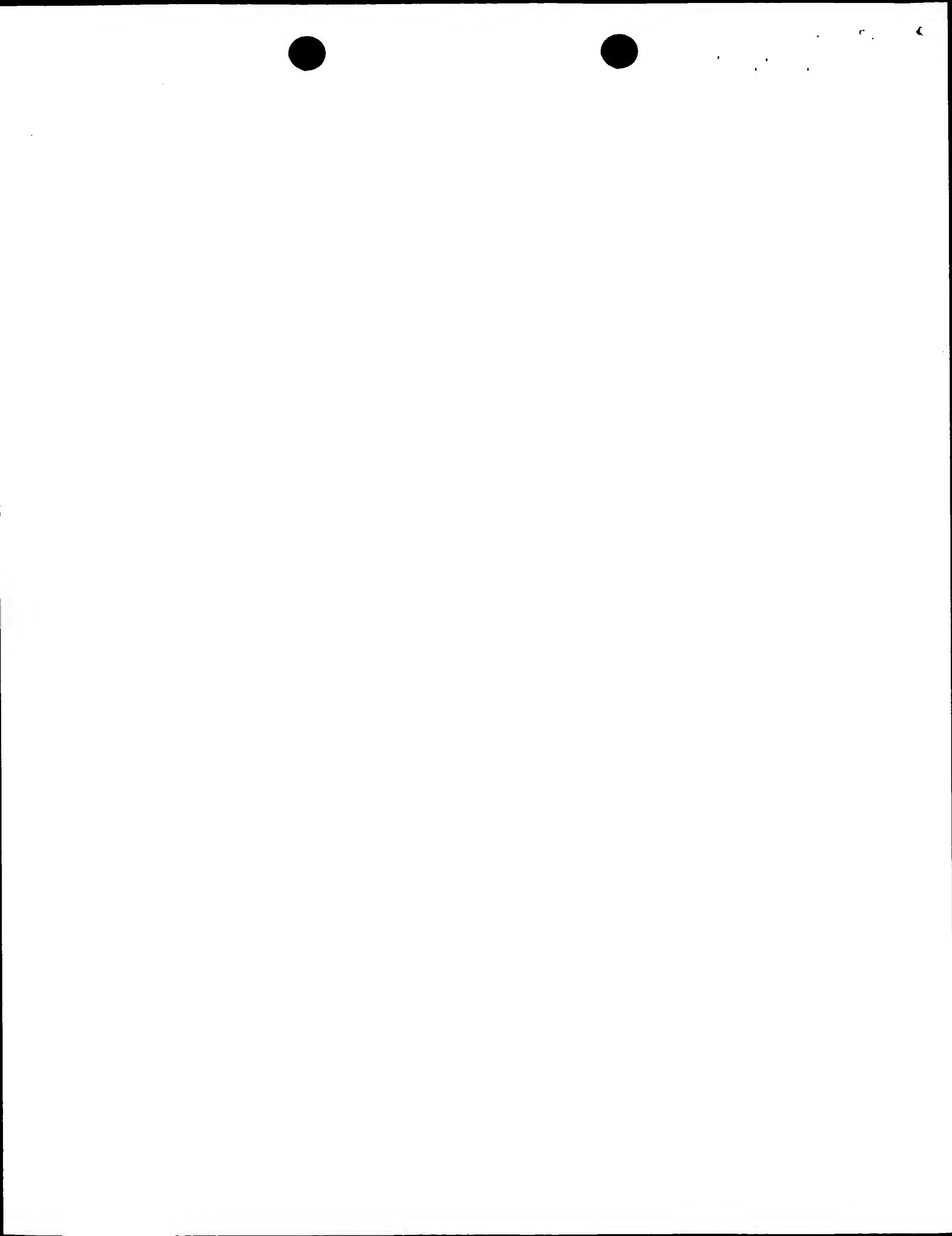
一方、近年、多数の蛋白質の立体構造が解析されており、その成果はプロテインデータバンクから利用できるようになっている。現在約 6,000 の蛋白質・核酸の構造が収録されており、機能の異なる独立の蛋白質は 400 程度である。これらの蛋白質の立体構造から、配列の相同性もなく進化的にも機能的にも全く関係がないように見える蛋白質が同じ構造モチーフを有する例が多数明らかにされている。

#### 発明の開示

本発明の第一の課題は、あるアミノ酸配列を有する蛋白質のアミノ酸配列情報から、その蛋白質がとる可能性の高いスキャッフォールドを推定することにより、立体構造をモデリングするための方法を提供することにある。正しいスキャッフォールドの推定は、立体構造を正しく必要な精度でモデリングするための出発点となり得る。すなわち、本発明の究極の課題は、アミノ酸配列の情報のみから蛋白質の立体構造を高い信頼性をもって推定する方法を提供することにあり、その手段としてスキャッフォールドを推定する方法を提供することが本発明の具体的課題である。また、本発明の別の課題は、上記の方法に有用なデータベースを提供することにある。

本発明者らは上記の課題を解決すべく銳意努力した結果、立体構造が既知または推定可能な蛋白質について各アミノ酸残基の側鎖の環境情報を備えたデータベースを作成し、そのデータベースを利用することによって、立体構造未知の蛋白質のアミノ酸配列の情報から、その蛋白質がとる可能性の高いスキャッフォールドを信頼性高く効率的に推定することができる方法を見出した。

すなわち本発明は、立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列



に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキヤッフォールドを推定する方法を提供するものである。スキヤッフォールドを推定した後に、鋳型蛋白質と質問配列の最適マッチングに基づいて質問配列に対応した立体構造（3次元座標）を得ることができる。

本発明の好ましい態様によれば、参照蛋白質のアミノ酸配列が該参照蛋白質の立体構造に基づいて連続した2以上のアミノ酸残基からなる2以上の部分配列に分割された上記方法；参照蛋白質のアミノ酸配列が疎水コアの形成に実質的に関与する1又は2以上のコア部分配列と疎水コアの形成に実質的に関与しない1又は2以上のサブ部分配列とに分割された上記方法；参照蛋白質における各アミノ酸残基の側鎖の蛋白質内部への埋没度及び／又は蛋白質表面への露出度の情報と、質問配列の各アミノ酸の疎水性及び／又は親水性の性質とに基づいてマッチングを行う上記方法；参照蛋白質の1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の片端又は両端においてはギャップを考慮せずにマッチングを行う上記方法；並びに、ギャップが1又は2以上のアミノ酸残基の削除又は付加である上記方法が提供される。

本発明のさらに好ましい態様によれば、マッチングが以下の工程：

- (a) 1又は2以上のコア部分配列を質問配列上でスライドさせながら、必要に応じて該コア部分配列の片端又は両端においてはギャップを考慮してマッチングを行う工程（ただし、2以上のコア部分配列を用いる場合には、該コア部分配列は参照蛋白質のアミノ酸配列での出現順に順番に配置する）；及び
- (b) 工程(a)でコア部分配列にマッチングされなかった質問配列の部分につき、該参照配列における各コア部分配列とのつながりにしたがって、必要に応じてギャップを考慮しつつサブ部分配列を質問配列上でスライドさせる工程を含む上記方法；参照蛋白質のアミノ酸残基の側鎖に関する環境情報と質問配列上の対応ア



ミノ酸残基の疎水度パラメータとから計算したスコアに基づいて最適マッチングを選択する上記方法；並びに、参照蛋白質についての自己マッチングスコアを用いて上記スコアを規格化する工程を含む上記方法が提供される。

別の観点からは、本発明により、立体構造が既知又は推定可能な1又は2以上の参照蛋白質のアミノ酸残基の側鎖についての環境情報を含むデータベースであって上記の各方法に用いられるためのデータベースが提供される。このデータベースは、通常の記憶媒体、例えば磁気ディスク、光ディスク、CD-ROM、磁気テープなどに格納された形態で流通可能であり、該環境情報と質問配列のアミノ酸残基の性質との一致度をスコアとするマッチングによって、質問配列から構成される蛋白質の安定なスキヤッフォールドを推定するために用いることができる。

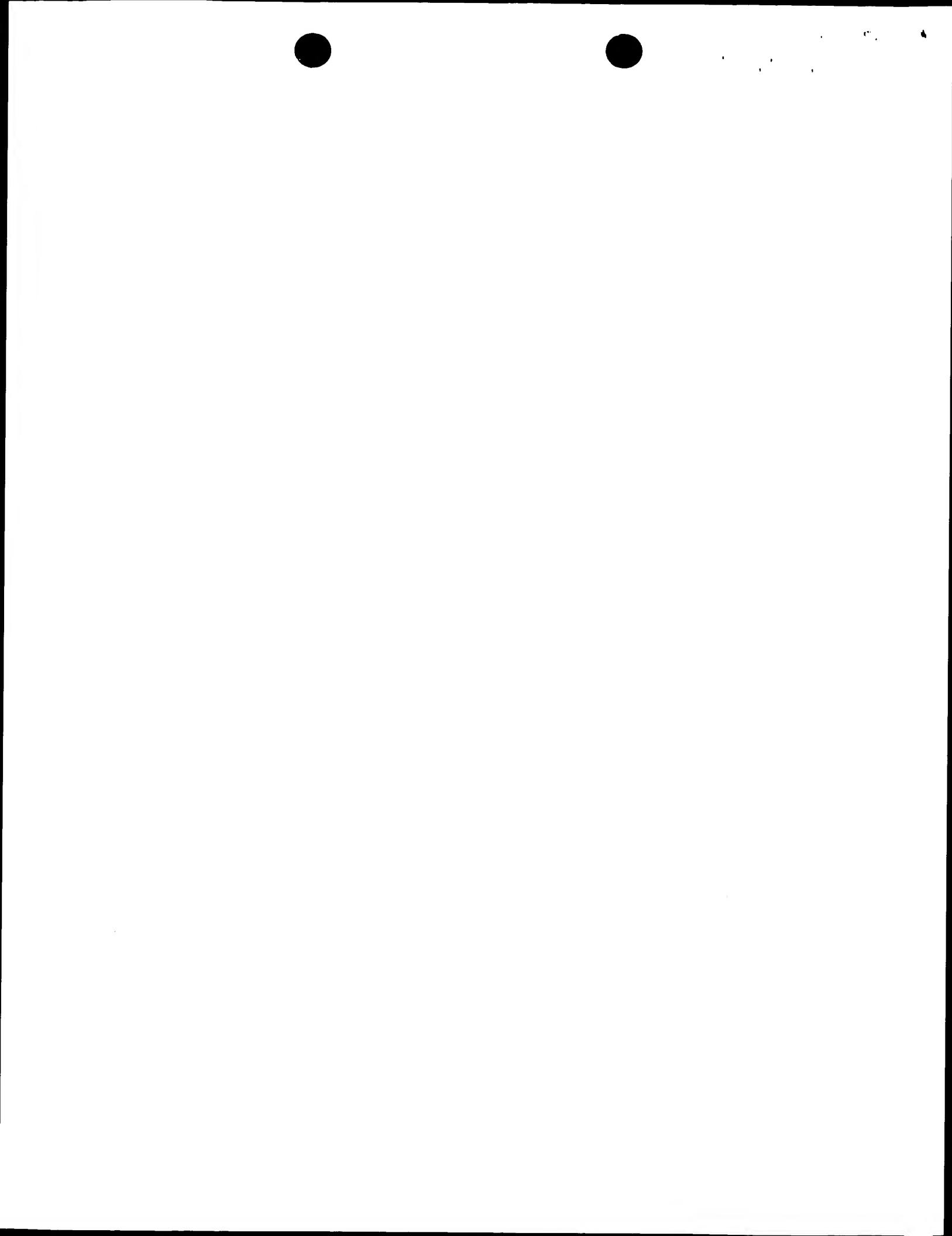
#### 図面の簡単な説明

第1図は、「立体構造」、「構造モチーフ」、及び「スキヤッフォールド」の関係を示した図である。図中、(a) は立体構造を示し、(b) は構造モチーフを示し、(c) はスキヤッフォールドを示す。

#### 発明を実施するための最良の形態

本発明の方法は、質問配列から理論的に可能な主鎖の折り畳み方を網羅・探索して評価するかわりに、参照蛋白質のデータベースから質問配列の蛋白質が立体構造的に類似する鋳型蛋白質を選び、鋳型蛋白質のスキヤッフォールドから質問配列の蛋白質の立体構造を推定することを特徴としている。本発明の方法は、一般的には、ワークステーション、パーソナルコンピューターなどの汎用のコンピューターを用いて高速に行うことができる。

本明細書において用いられる用語は、以下に述べる概念を含めて、最も広義に解釈する必要がある。「立体構造」とは3次元座標で表される蛋白質の構造を意味しており、アミノ酸残基の側鎖など存在する全原子を含む場合もあるが、それらの一部を省略することもある。「モデリング」とは、立体構造が実験的に解明

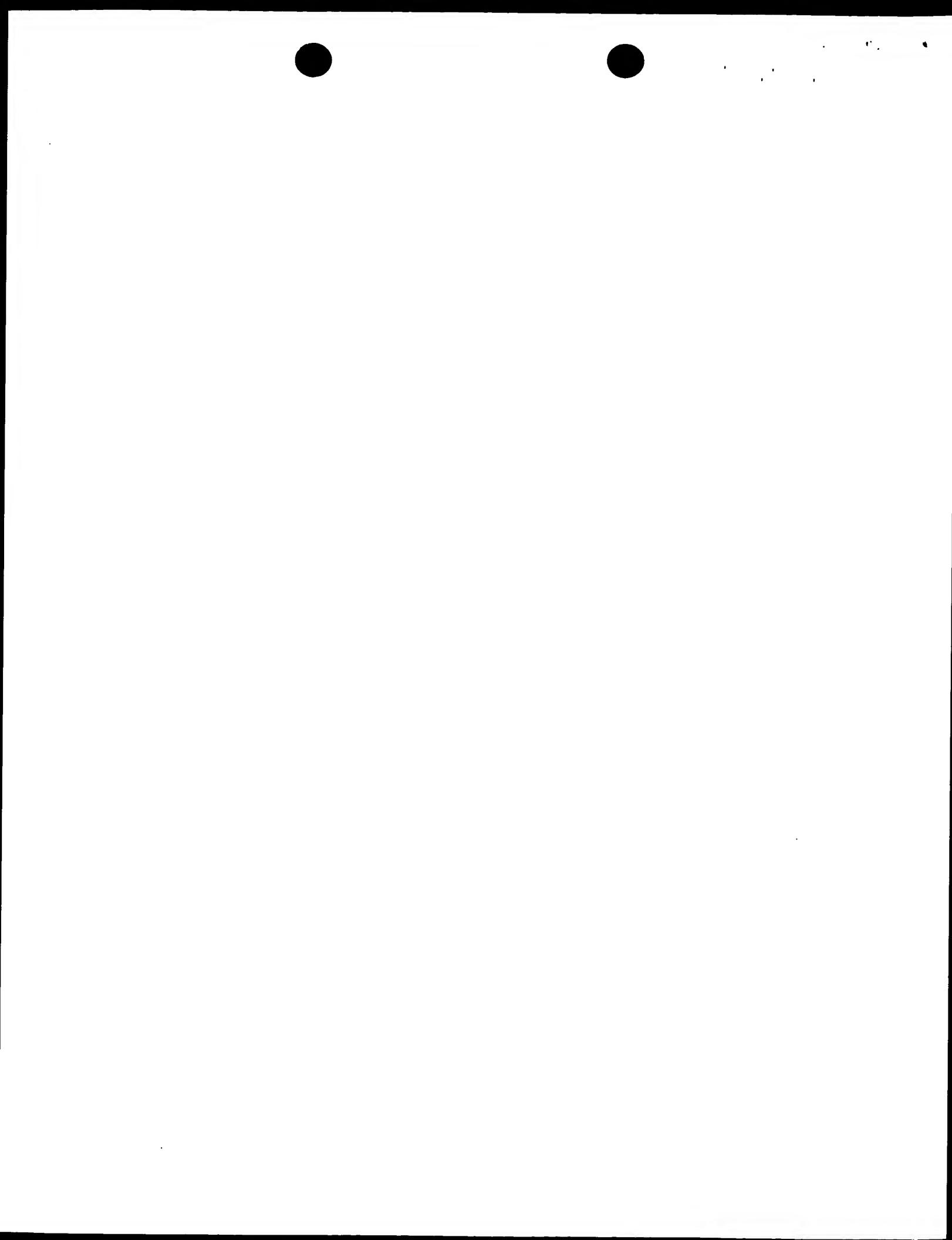


されているか否かに係わらず、ある蛋白質について存在可能性の高い立体構造を構築して原子の3次元座標で表現することをいう。

「2次構造」、「構造モチーフ」、及び「スキヤッフォールド」などの用語は Branden 及び Tooze らの著書に記載されている (Carl Branden and John Tooze, *Introduction to Protein Structure*, Garland publishing Inc. New York, 1991: 和訳「タンパク質の構造入門」、教育社、1992)。「構造モチーフ」と「スキヤッフォールド」は、ともにペプチド主鎖のみのトポロジーを表す点では共通しているが、「構造モチーフ」が2次構造群の組み合わせとペプチド鎖の流れを平面的かつ模式的に表現するのに対して、「スキヤッフォールド」は2次構造群の集合状態を含む蛋白質の3次元構造の骨組を意味する。「立体構造」、「構造モチーフ」、及び「スキヤッフォールド」の関係を図1に示す。

立体構造を推定したいアミノ酸配列を「質問配列」と呼び、その質問配列から構成される蛋白質を「質問配列の蛋白質」と呼ぶ。立体構造が既知または推定可能な蛋白質であって本発明のデータベースに含まれる蛋白質を「参照蛋白質」と呼び、参照蛋白質のうち、質問配列とのマッチングスコアがよく、質問配列の蛋白質と立体構造の類似性が高い蛋白質として選別された参照蛋白質を「鋳型蛋白質」という（「鋳型候補蛋白質」と呼ぶ場合もある）。鋳型蛋白質のスキヤッフォールドは、質問配列の蛋白質の立体構造を構築する際の鋳型として用いられる。

一般に2以上の配列をアミノ酸残基の一致度や相同性が高くなるように対応づけて並置する操作又は結果は「アラインメント」と呼ばれるが（「一致度」とは配列間で対応する残基間の厳密な一致を意味しており、「相同性」とは同等または類似など曖昧さを含めた一致の割合を意味する）、本発明の方法において「マッチング」（又は「対応付け」）という場合には、アミノ酸残基の一致又は相同性によらず、環境情報と性質の一致で残基を対応づけて並置する操作又は結果を意味している。ある対応付けにおける参照蛋白質の各アミノ酸残基の環境と質問配列のアミノ酸残基の性質との一致度を「マッチングスコア」（又は単に「スコア」）とよぶ。「環境情報」は、主として、参照蛋白質の立体構造における各アミノ酸

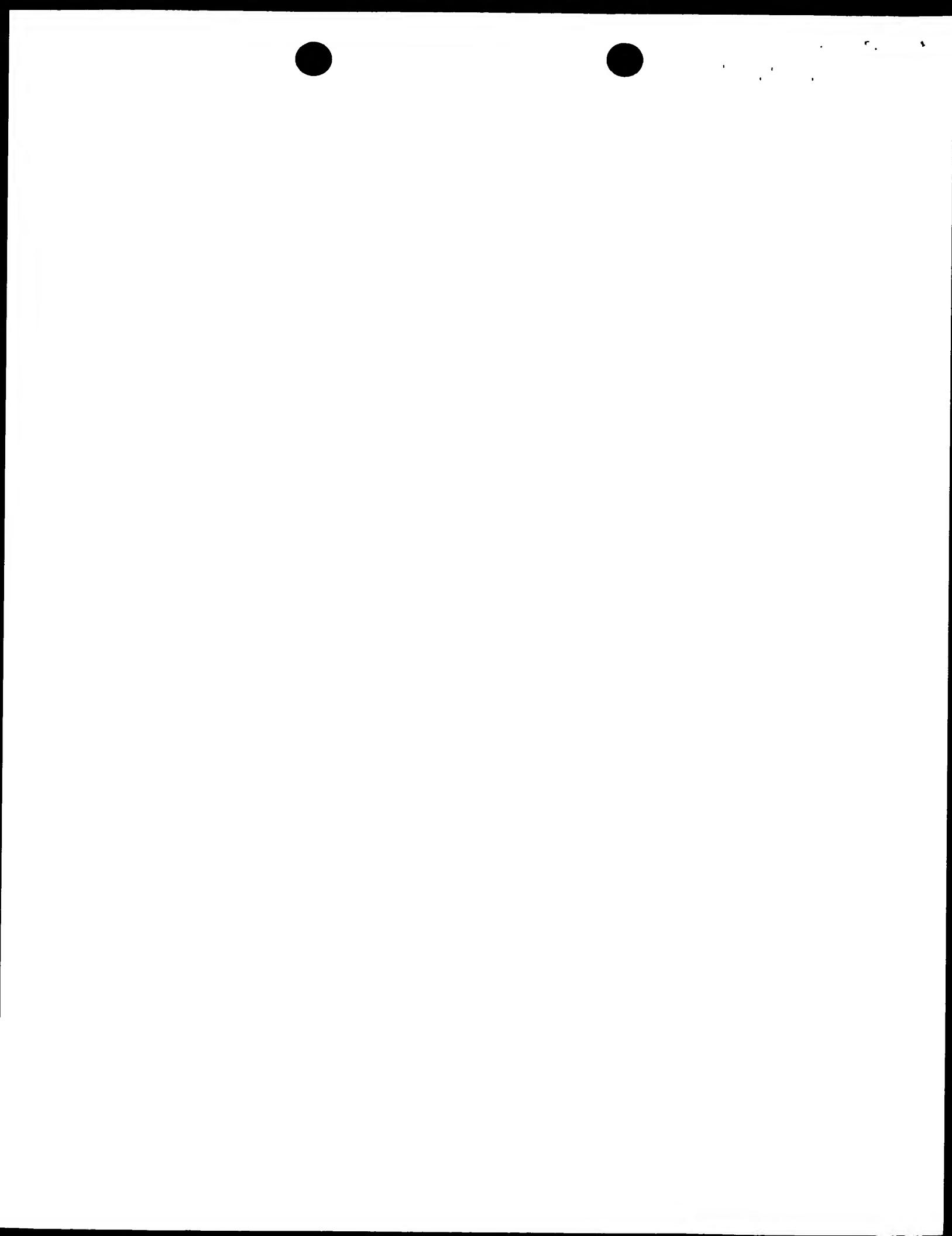


残基の側鎖の蛋白質表面への露出度や存在環境を示す。アミノ酸配列について用いられる「ギャップ」という用語は、2以上のアミノ酸配列間の対応付けにおいていずれか一方の配列に対応するアミノ酸残基がない部分を指し、一方の配列から見ると1又は2個以上のアミノ酸残基の挿入及び／又は欠失していることを意味する。

本発明の好ましい態様では、このデータベースに参照蛋白質毎に立体構造を反映して分割した2以上の部分配列の情報とアミノ酸残基毎の環境情報とを格納しておく。参照蛋白質の環境情報と質問配列中の対応アミノ酸残基の疎水度パラメータとから計算されるマッチングスコアに基づいて2つの配列をマッチングする。疎水度パラメータは20種のアミノ酸残基に対して予め数値化しておくことが望ましい。

部分配列のうち、疎水コアの形成に関与するコア部分配列についてはギャップを入れずに質問配列上をスライドさせ、両端においてのみアミノ酸残基の増減(隣接するサブ部分配列の末端から1又は2個以上のアミノ酸残基をとって該コア部分配列に組み込むか、又はその逆の操作を意味する)を考慮してマッチングを行い、スコアのよい鋳型候補蛋白質を選別する。次に、疎水コアの安定化に関与しないサブ部分配列について必要に応じてギャップを考慮したマッチングを行い、鋳型候補の数を絞る。最終的な鋳型蛋白質の選択は、各蛋白質の最適マッチングのスコアを自己マッチングスコアで規格化して比べることにより行うことができる。

本発明の方法は、マッチングスコアが高いほど2つの配列間でスキヤッフオーラドや立体構造の類似性が高いとの仮定に基づいており、配列の相同意が殆どない配列間の残基のマッチングを蛋白質の折り畳み原理に従って行うことによって、参照蛋白質からの鋳型蛋白質の適切な選択を可能にしている点に特徴がある。ある。本発明の好ましい態様では、(1) データベースの作成；(2) 部分配列を用いたマッチング；(3) マッチングスコアの計算；(4) 参照蛋白質から鋳型候補蛋白質の選択；(5) 鋳型蛋白質の選択を主要な要素としている。



一般的な水溶性の単一サブユニットの蛋白質を構成するペプチド鎖については、アミノ酸残基の疎水性側鎖ができるだけ分子内部に埋もれて露出せず、一方、親水性側鎖はできるだけ蛋白質分子表面に露出するのが自由エネルギー的に有利である。もっとも、細胞膜に相互作用する蛋白質や複数のサブユニットからなる蛋白質など存在環境が異なる蛋白質では、個々の蛋白質構造やサブユニット構造だけをみると、疎水性側鎖が蛋白質表面に露出していることもある。本発明の方法では、このような蛋白質の立体構造の多様性も考慮されており、個々のアミノ酸残基の存在環境を反映したスコア付けができるようになっている。アミノ酸残基毎の特定2次構造のとり易さの概念（例えば P. Y. Chou, & G. D. Fasman, *Adv. enzymol.* 47, 45, 1978）は基本的には用いないが、必要に応じてそれらの概念を加味したスコアを用いることができる。

### 実施例

以下、本発明の方法の好ましい態様をスキームで示し、このスキームに従って本発明の方法をより具体的に説明するが、本発明の方法はこのスキームの細部又は以下の説明の細部に限定されることはない。



立体構造が既知または推定可能な蛋白質  
(参照蛋白質) のアミノ酸配列と三次元座標

コア部分配列とサブ部分配列への分割  
アミノ酸毎に環境情報アサイン  
(露出度パラメータと存在環境)  
自己マッチングスコア計算  
その他の情報アサイン

部分配列情報、アミノ酸毎のパラメータ ←  
含む参照蛋白質のデータベース

質問配列 ←→ 参照蛋白質を1個とりだす

コア部分配列でのマッチング  
各配列の両端での残基の増減

全コア部分配列の最適マッチング

全参照蛋白質間でスコアの比較

鋳型候補蛋白質 (数個～数十個)  
(推定スキャッフォールド)

サブ部分配列のマッチング

鋳型蛋白質 (1～数個)  
(推定スキャッフォールド)

側鎖の置換、ギャップの処理

立体構造の構築 (モデリング)

構造最適化  
エネルギーの比較、機能面の考察

立体構造の推定



### (A) データベースの作成

立体構造が既知または推定可能な参照蛋白質について、立体構造に関わる情報として、各アミノ酸残基の環境情報及び部分配列の情報を含むデータベースを作成しておく。データベースに収納する蛋白質としては、立体構造の情報が既知または推定可能な蛋白質であれば、すべてエントリーとすることができます。同時に構造決定された蛋白質中に複数のサブユニットが含まれるときは、独立のエントリーとしておくことが好ましい。ペプチド鎖で繋がった複数ドメインからなる構造については、全体構造とともに各ドメインも独立エントリーとしておくことが好ましい。

参照蛋白質毎に含める情報は以下の通りである。

#### (1) 一般的な情報

- (a) 蛋白質名（蛋白質コード）、サブタイプ、アミノ酸数、アミノ酸配列、ドメイン、サブユニットなどに関する情報。
- (b) 立体構造に関する情報として、立体構造の決定（又は推定）方法（結晶解析法、n m r 法、又はモデリング法のいずれにより立体構造を決定したか）、P D B コード、モデリング法による場合には鋳型として利用した蛋白質分子名、結晶解析の場合には共結晶化された分子がある場合にはその化学名などの情報のうち適宜のものを含めておく。
- (c) 生物学的意義に関する情報として、生物学的機能、生物種、存在する組織・臓器、エフェクター分子などの情報を含めておく。

#### (2) 部分配列情報

蛋白質の立体構造における特徴に従って、配列を分割して2以上の部分配列とし、各部分配列について以下の情報をデータベースに含めるのが望ましい。

- (a) N末端から何番目の部分配列か
- (b) コア部分配列かサブ部分配列かのフラグ
- (c) 始点と終点（N末端からのアミノ酸配列番号で）、配列の長さ、その距離及



びベクトル、サブ部分配列ならば最短の残基数

(d) 部分配列間の距離及びベクトル

(e)  $\beta$ -シート形成の水素結合をする相手の部分配列番号、および逆平行または平行  $\beta$ -シートなどの区別、あるいは一定の距離内にある部分配列番号など

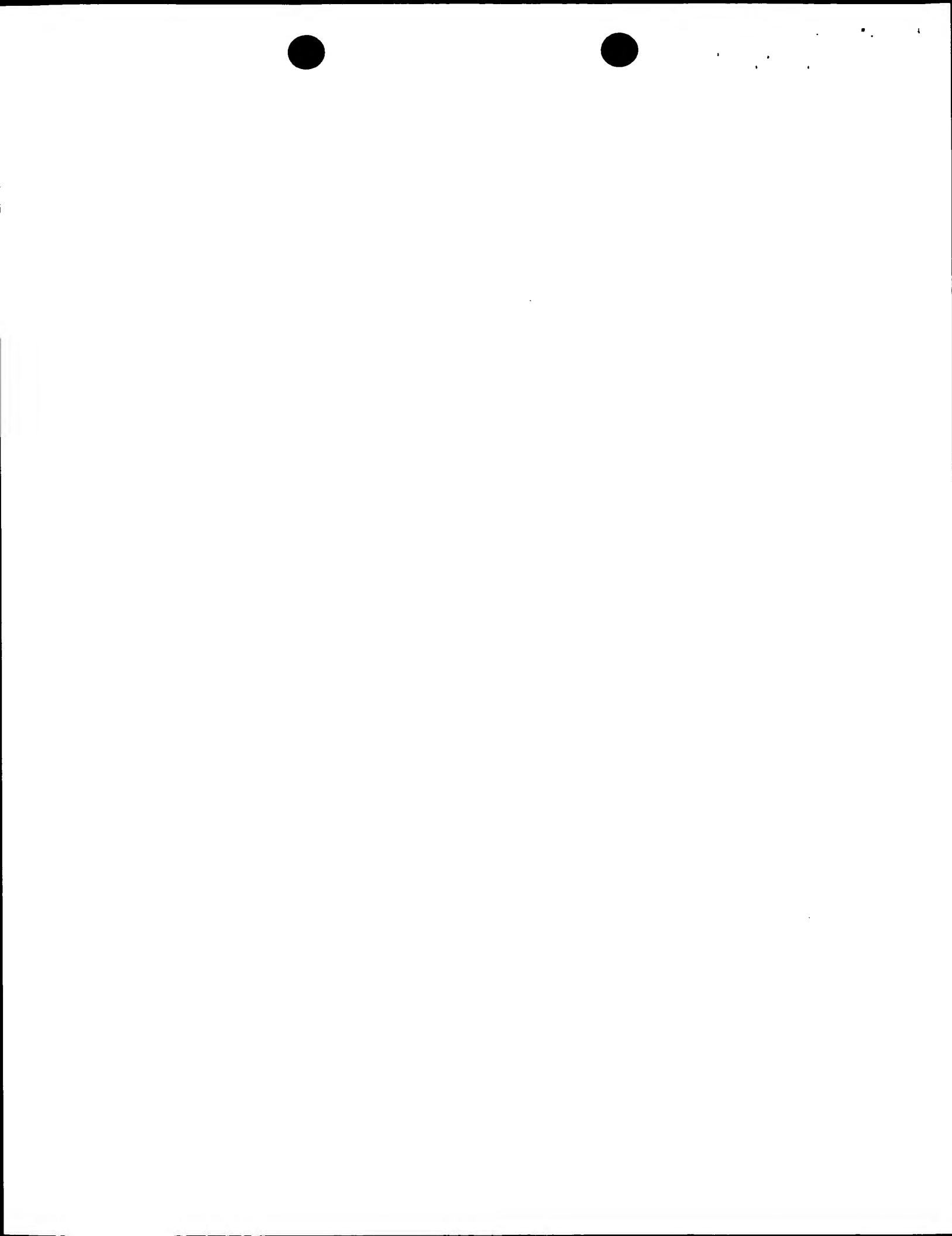
(f) 自己マッチングスコア (後述)

部分配列への分割の基準は特に限定されないが、 $\alpha$ -ヘリックス構造や  $\beta$ -ストランド構造などの2次構造を形成し実質的に疎水コアの形成に関わる部分をコア部分配列とし、それ以外の部分をサブ部分配列とすることができる。各部分配列は、例えば、7残基以上を含む連続した配列とすることが望ましい。 $\beta$ -ターン構造については、はじめからコアまたはサブ部分配列に加えて構わないし、別のフラグによって区別し、検索時にコアかサブかの判断を加えてもよい。分割の作業は、コンピュータグラフィックス画面上で対話的に個々の蛋白質について行ってもよいし、分割の基準を定めたプログラムを作成して自動的に行ってもよい。その判断基準として、例えば、隣り合う4つのCアルファ原子のなす擬ねじれ角などの数値を基準に用いれば、自動的に部分配列に分割できる。

### (3) アミノ酸残基毎の環境情報

#### (a) 露出度パラメータ

参照蛋白質の全てについて、立体構造に基づいて残基毎の蛋白質表面への側鎖の露出度及び蛋白質内部への側鎖の埋没度を計算し、その計算値に基づいて露出度パラメータを割り振る。本明細書において用いられる「露出度パラメータ」という用語は、各残基の側鎖がどの程度分子表面に露出しているか、又は埋没しているかを示す数値を意味している。露出度パラメータを定義する方法は特に限定されず、いかなる方法を採用してもよいが、露出度が高いものはマイナスの値、露出度が低いものはプラスの値をとるように設定するのが望ましい。例えば、立体構造において各アミノ酸側鎖の溶媒への接触可能表面と蛋白質原子への接触表面積を計算して、その差から露出度パラメータを算出して利用することができる。



また、例えば、全分子表面に対する溶媒接触表面の割合を基準として定めることもできる。

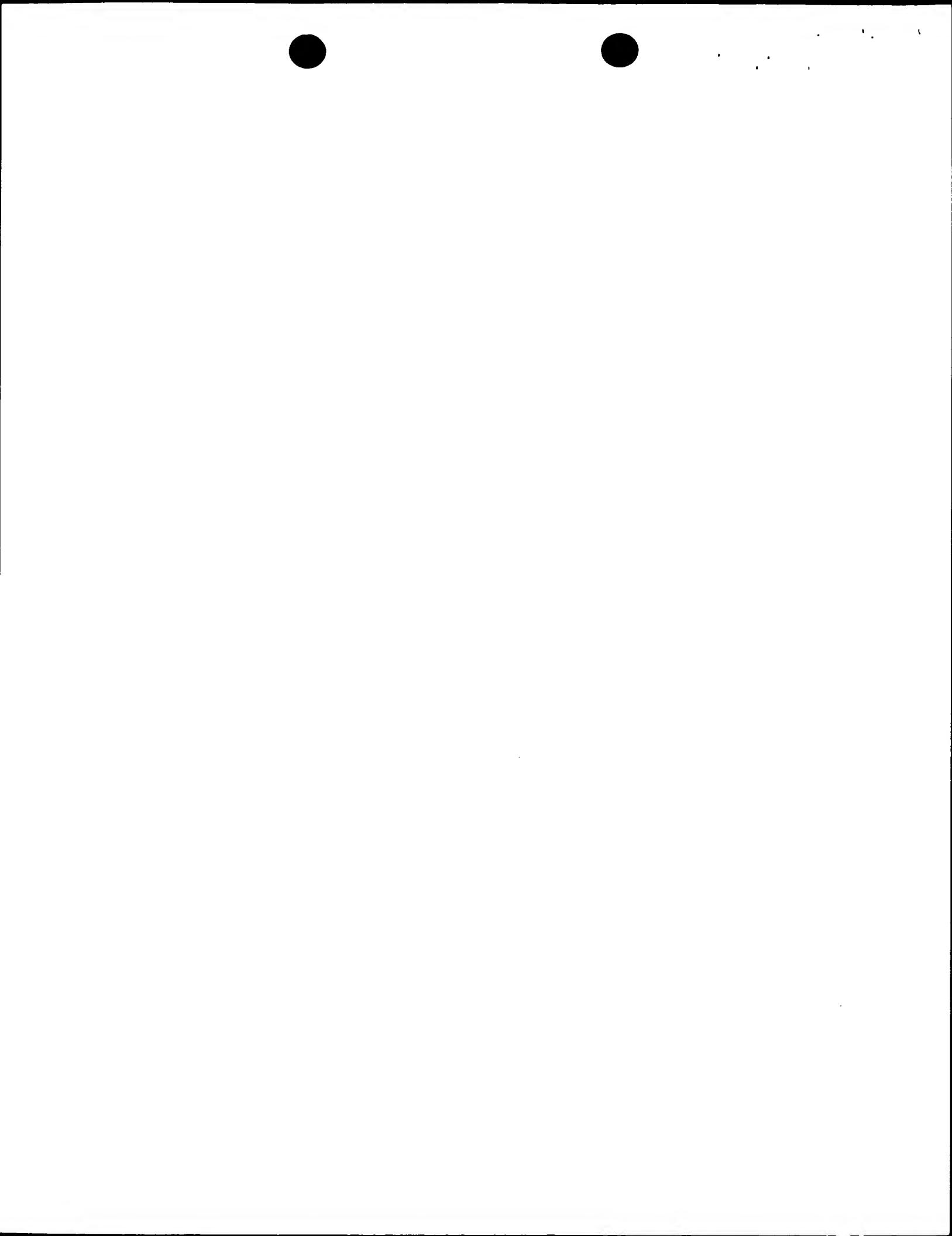
(b) 存在環境フラグの設定

細胞膜に相互作用する蛋白質では膜と相互作用する部分の表面に疎水性アミノ酸残基の側鎖が露出しており、また、サブユニットやドメイン構造が集合して安定化する接触面を有する蛋白質では、接触面に疎水性アミノ酸残基の側鎖が露出している場合があり、これらの蛋白質は単独で存在する水溶性蛋白質とは異なる性質を有している。このような蛋白質については、一般的には、先述の折り畳み原理はそのままでは適用できない。データベース作成時に、例えば、次のような存在環境フラグを露出度パラメータとは別に各アミノ酸残基に与えておくことができる。

例えば、蛋白質の由来や機能に関する実験結果および立体構造から、各アミノ酸の存在環境が以下のいずれに相当するかを推定して環境フラグを与えておき、マッチングやスコアの計算の際に考慮することが可能である。

- 0 : 不明 (未定義または定義できず)
- 1 : 分子内部 (蛋白質内、サブユニット内、ドメイン内の接触)
- 2 : 分子内孔 (リガンド結合部位)
- 3 : 分子表面 (水環境と接触)
- 4 : 分子表面 (別蛋白質、別サブユニット、別ドメインと接触)
- 5 : 分子表面 (膜と接触)

さらに、立体構造形成に特殊な影響を与えるアミノ酸残基については、特殊残基であることを示すフラグを与えておき、マッチングやスコアの計算の際に考慮することができる。例えば、S-S結合しているシステイン残基やプロリンのように主鎖に水素結合性官能基が不足しているアミノ酸残基、または側鎖原子間に強い親水性相互作用を形成する可能性のある残基などに適用することができる。



### (B) 質問配列と参照蛋白質の配列間のマッチング

一方の配列を他方の配列上スライドさせて最適のマッチングを効率よく探すために、部分配列の概念を利用することができる。そのために、上記データベースは、各参照蛋白質について立体構造から得られる部分配列の情報とアミノ酸残基ごとの環境情報がアミノ酸配列順に収めてある。一方、質問配列が有している情報はアミノ酸配列の情報だけであり、アミノ酸残基毎の疎水度パラメータの表からとった値を当てはめてスコアの計算に用いる。データベースから参照蛋白質を1つずつ取り出し、アミノ酸配列中に出現する順に並べた部分配列群を質問配列上でスライドさせて、部分配列群と質問配列との間でマッチングスコアの最もよいマッチングを探す。

#### (1) 部分配列を用いたマッチング

部分配列を用い、かつコア部分配列ではギャップを考慮しないでアミノ酸残基単位でマッチングを行うことにより、最適マッチングの配列間対応付けを高速に検索することができる。一般に、進化の過程では残基の置換とともに挿入や欠失が起きることが多く、その考慮は配列のマッチングに不可欠である（「従来の技術」の欄を参照）。しかしながら、一般的には、ギャップが入るのはサブ部分配列であることが多い。これは、疎水コアの安定化に関するコア部分配列でその両端以外で挿入や欠失が起きると、その安定スキヤッフオールド自体が損なわれ、蛋白質の立体構造が大きく変化してしまうからである。

そこで、本発明の方法の好ましい態様では、コア部分配列とサブ部分配列とに分けて2段階のマッチングを行う。第一段階のコア部分配列を用いたマッチングにおいては、まずギャップを考慮せずに1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の両端においてのみアミノ酸残基の増減を考慮したマッチングを行って最適マッチングを探索する。

各コア部分配列について質問配列の上をスライドさせながらマッチングスコアを計算して保存する。全部のコア部分配列について同様に計算した後、全体とし



て最適なマッチングを決定する。2以上のコア部分配列を用いる場合には、質問配列上に2以上のコア部分配列を参照蛋白質のアミノ酸配列中での出現順序に従って重なり合わないように配置し、コア部分配列間には4個程度以上のアミノ酸残基の存在を仮定して（立体的に結合可能なアミノ酸残基数で隣のコア部分配列と順番に繋がるという条件、例えば $\beta$ -ターンなどに要するアミノ酸残基数）、それぞれのコア部分配列の順序を変えずに質問配列上をスライドさせ、最も総スコアのよいマッチングを選択する。この際、各コア部分配列のマッチングスコアが最大である必要はない。この第一段階で、高いスコアを与える鋳型蛋白質を数個から数十個選択し、スキヤツフォールド候補として第二段階に進む。

#### (C) マッチングスコア

マッチングスコアは参照蛋白質の環境情報の露出度パラメータ EP(i) と質問配列の残基の疎水度パラメータ HB(j) を用いて計算する。i は参照蛋白質のアミノ酸配列中のアミノ酸残基番号であり、j はそれと対応づけられた質問配列の残基番号を指す。マッチングスコアの計算式は、参照蛋白質の分子内部に埋没した側鎖環境に質問配列の疎水性の強いアミノ酸残基が対応し、分子表面に露出した側鎖環境に質問配列の親水性の強いアミノ酸残基が対応すると高いスコアが得られるような計算式であれば、いかなるものを利用してもよい。マッチングスコアは例えば、最も単純には次式によって計算することができる。

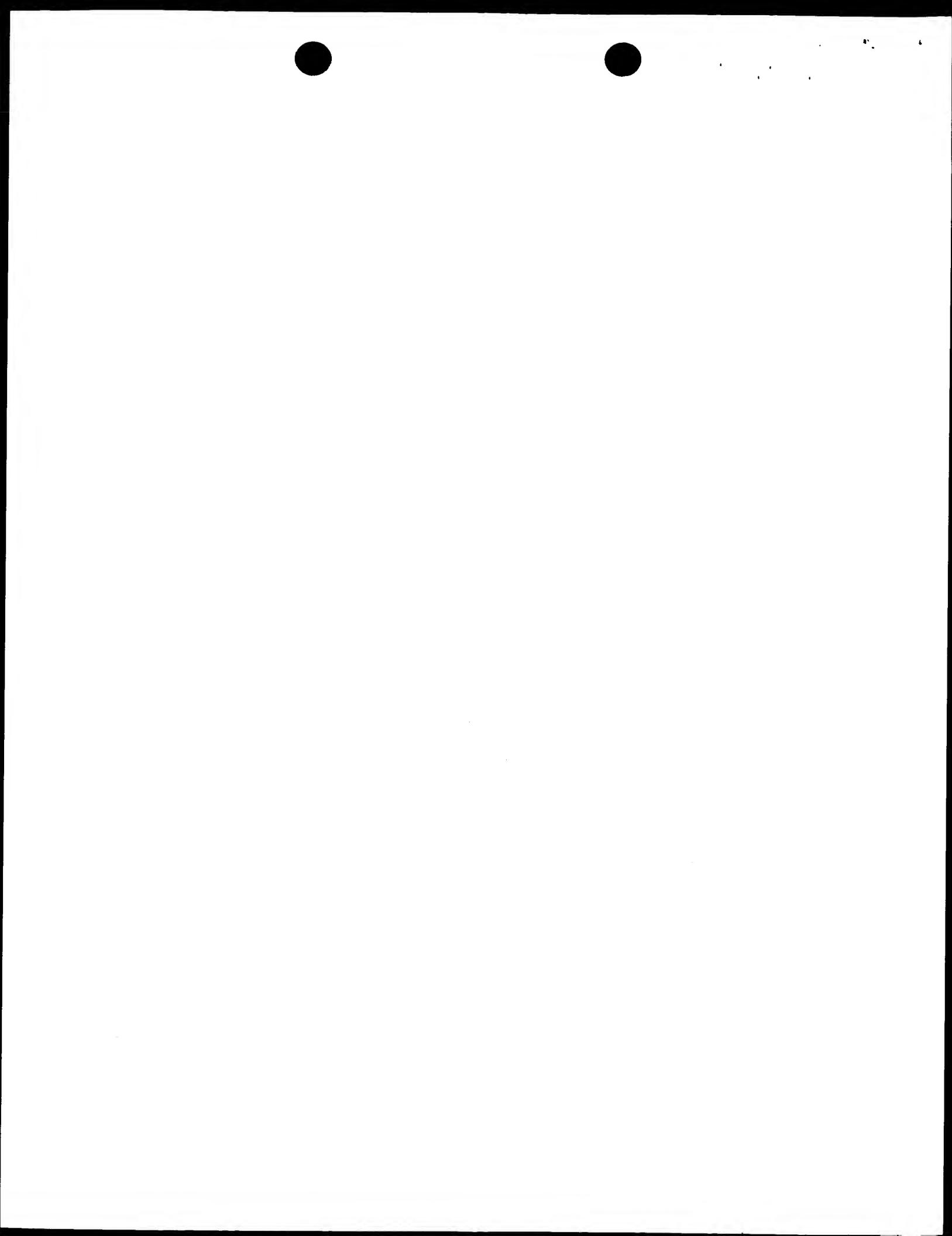
$$\text{残基毎のマッチングスコア} = EP(i) \times HB(j)$$

部分配列のマッチングスコア = 配列に含まれる残基のマッチングスコアの和

全配列のマッチングスコア = 全部分配列のマッチングスコアの和

#### (1) 疎水度パラメータ

20種のアミノ酸残基のそれぞれに疎水性又は親水性の性質に関連した疎水度パラメータを与えておく。疎水度パラメータの決定方法は特に限定されず、いかな



る基準による値を用いてもよい。例えば、文献に記載されたアミノ酸毎の疎水性値を用いてもよく、又は適宜の方法により独自の基準で算出したものを用いてもよい。また、あるアミノ酸について、結晶解析された蛋白質中の全出現回数に対して蛋白質分子内部に埋没された残基の比率を統計的に求めておき、その比率を該アミノ酸の疎水度パラメータとして利用してもよい。

また、例えば、個々のアミノ酸残基に別々の値を与えてよいが、次のように段階化した疎水度パラメータを与えることもできる。

表 1

2 : 強い疎水性 (イソロイシン、バリン、ロイシン、フェニルアラニン)

1 : 弱い疎水性 (アラニン、メチオニン、シスチン、チロシン)

0 : ほぼ中性 (グリシン、プロリン、リジン、アルギニン)

-1 : 弱い親水性 (スレオニン、ヒスチジン)

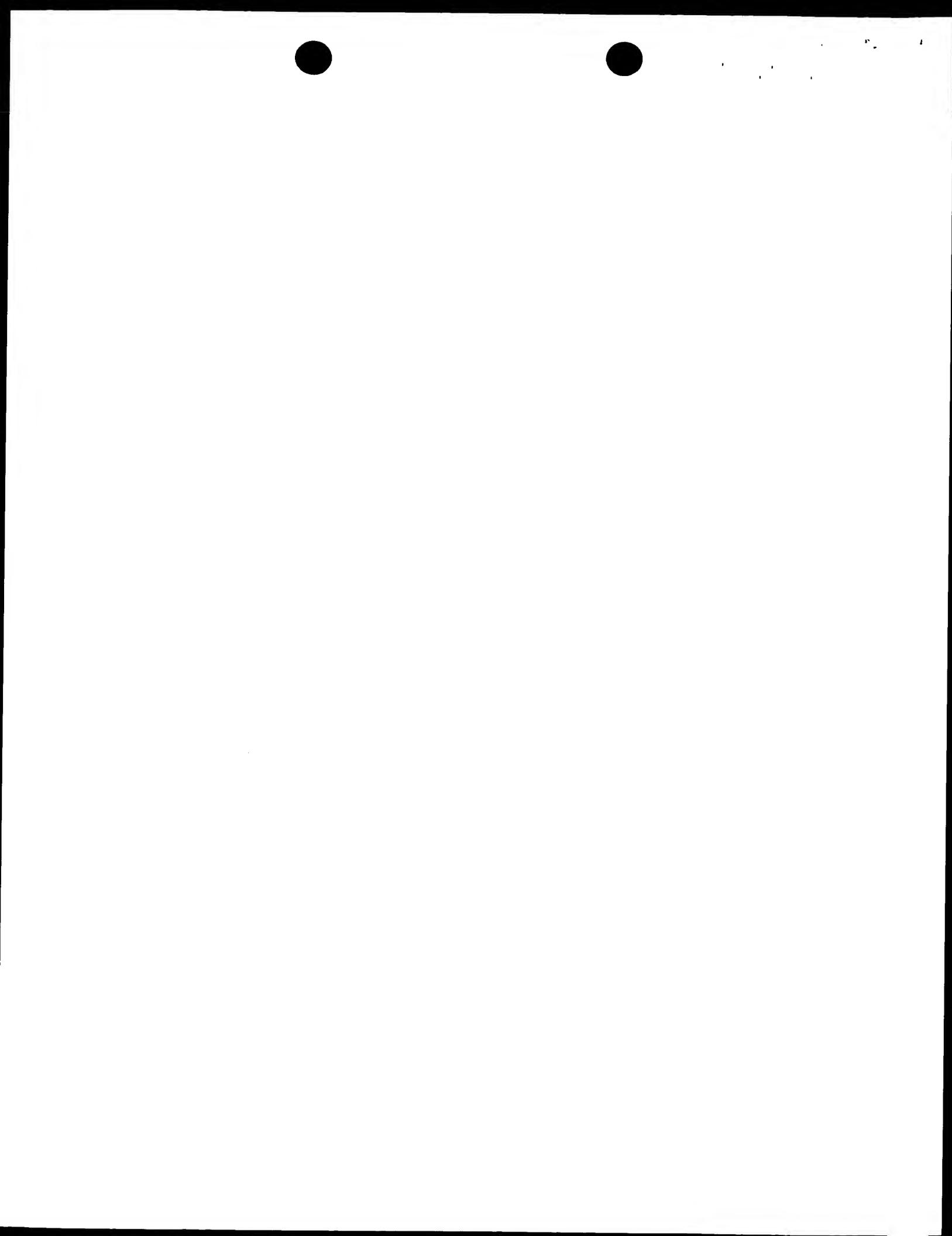
-2 : 強い親水性 (セリン、アスパラギン、アスパラギン酸、グルタミン、グルタミン酸)

## (2) 自己マッチングスコア

アミノ酸数及びアミノ酸組成の異なる蛋白質間で質問配列へのマッチングの良さを比較するためにはスコアの規格化をしておくことが望ましい。そのために、各参照蛋白質について、それ自体のアミノ酸配列の露出度パラメータ EP(i) とデータベースに用意された環境情報の疎水度パラメータ HB(i) とからマッチングスコアを予め計算してデータベースに保存しておく。例えば、次式のように計算するとよい。

$$\text{自己マッチングスコア} = \sum (EP(i) \times HB(i))$$

全部分配列の質問配列へのマッチングが終了して最適マッチングが得られた後、



得られたマッチングスコアに自己マッチングスコアを乗じて規格化することができる。全参照蛋白質について規格化された最適マッチングスコア同士を比較して最適な鋳型候補蛋白質を選択することができる。自己マッチングスコア及びマッチングスコアは、アミノ酸残基数が多いほど大きな値を取りやすい。

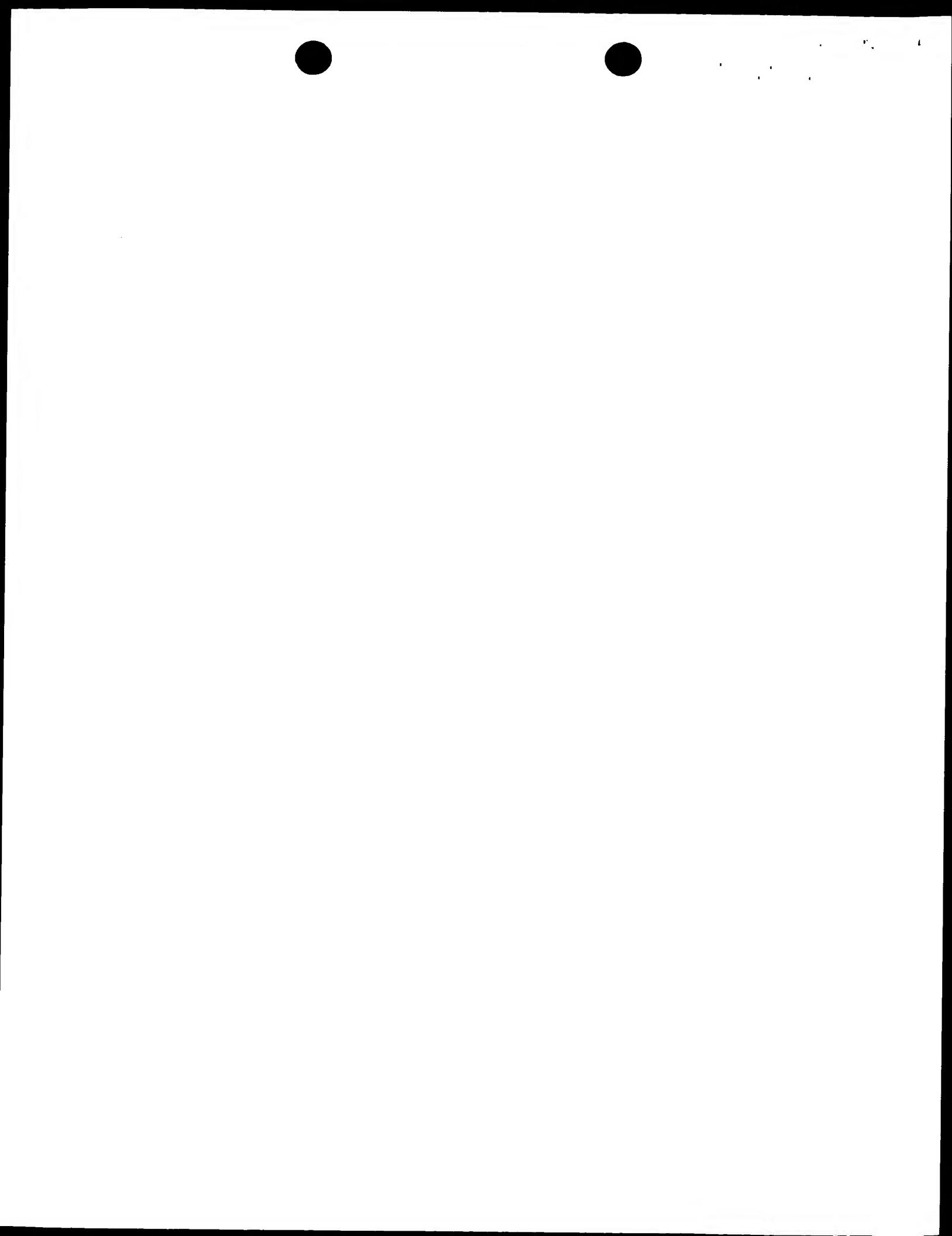
#### (D) 鋳型候補蛋白質の選択

鋳型候補蛋白質の選択の手順は概ね以下の通りである。

- (a) データベースから参照蛋白質を1個づつ取り出し、質問配列に対しマッチングを行う；
- (b) コア部分配列につきギャップを考慮せずに質問配列上をスライドしマッチングスコアを算出する；
- (c) 必要に応じて、各コア部分配列のN末端またはC末端の残基を増減しながらマッチングし、最適マッチングを得る；
- (d) 参照蛋白質全部について工程(a)～(c)を行い、最適マッチングとマッチングスコアを得る；
- (e) 規格化したスコアにより参照蛋白質からスコアのよいものを鋳型候補蛋白質とする（この段階で、それらの構造は質問配列のスキヤッフォールドの候補とすることができる）；
- (f) コア部分配列間をつなぐサブ部分配列のマッチングを行う。質問配列の対応する配列部分との配列長の違いやギャップの存在を考慮し、最適のマッチングとマッチングスコアを得る；及び
- (g) 規格化したスコアによって鋳型蛋白質を選択する。

#### (E) 立体構造の構築

質問配列の蛋白質の立体構造は、鋳型蛋白質のスキヤッフォールドの構造、及び該参照蛋白質と質問配列との最適マッチングの結果に基づいて、アミノ酸残基の側鎖の置換を行うことにより構築することができ、質問配列に対応した3次元座

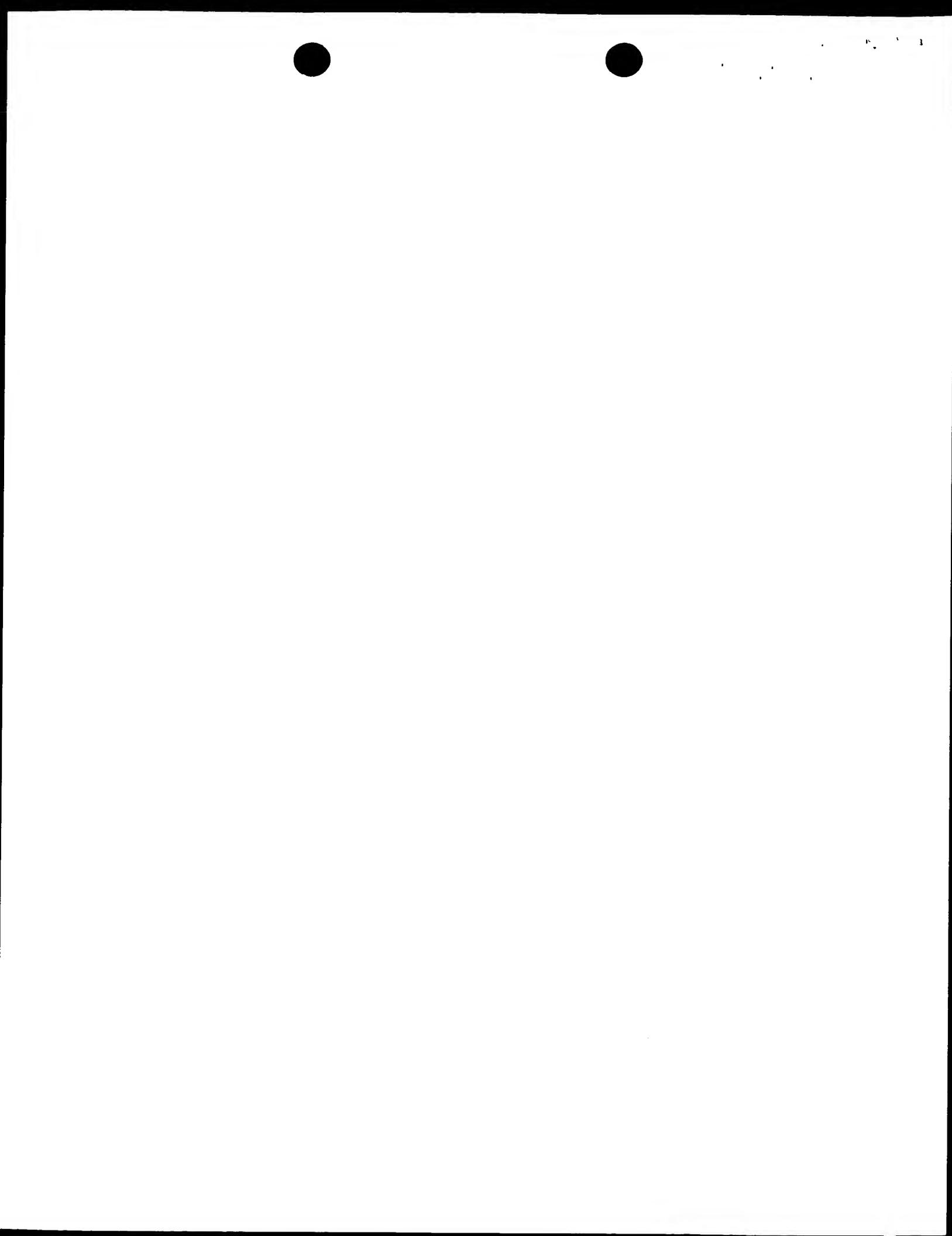


標を得ることができる。優劣をつけ難い2以上の鋳型蛋白質がある場合には、すべてについて立体構造を構築することが望ましい。サブ部分配列の長さが鋳型候補と異なる場合には、結晶構造に現れたループ構造を集めたデータベース等を用いて、該サブ部分配列の適切なトポロジーを決定することができる。鋳型のスキヤッフォールドが修正される部分については、マッチングスコアと同様に残基の性質と露出度を考慮しつつトポロジーを決定することができる。重大な立体障害や立体構造を不安定化する分子内部の隙間などの有無を確認し、また構造最適化計算や分子動力学計算により構造の微調整を行ったのちに立体構造の安定性を比較する。

最後に、全ての鋳型に基づいて構築された立体構造をエネルギーとマッチングスコアにより順位付けするが、該質問配列の蛋白質の機能が既知の場合には、該鋳型蛋白質に知られた機能との対応、その機能の発現に関与するものと推定されるアミノ酸残基の立体構造上の位置の妥当性、さらにはアミノ酸変異による機能への影響などの情報を鋳型蛋白質の選別に利用できる。

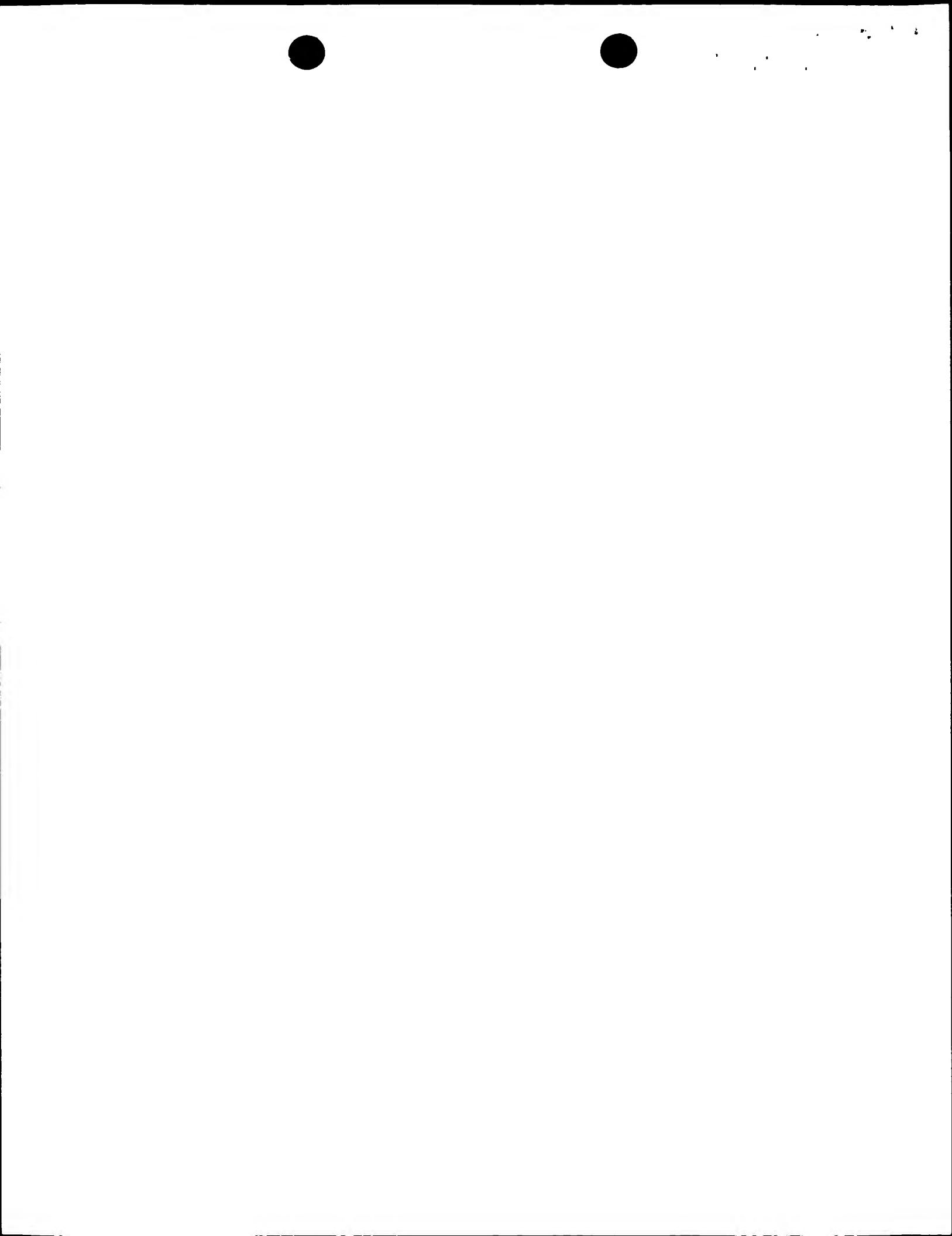
#### 産業上の利用可能性

本発明の方法によれば、立体構造が既知または推定可能な蛋白質のアミノ酸配列データベースに基づいて、アミノ酸配列の情報のみからそのアミノ酸配列により構築される蛋白質の立体構造に関する情報を信頼性高く効率的に入手することができる。



## 請求の範囲

1. 立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースを用い、参照蛋白質の各アミノ酸残基の環境情報と質問配列中の各アミノ酸残基の側鎖の疎水性又は親水性の性質とに基づいてマッチングを行い、参照蛋白質の中から質問配列の蛋白質と立体構造の類似性が高い鋳型蛋白質を選択して質問配列の蛋白質のスキップオールドを推定する方法。
2. 参照蛋白質のアミノ酸配列が該参照蛋白質の立体構造に基づいて連続した2以上のアミノ酸残基からなる2以上の部分配列に分割された請求の範囲第1項に記載の方法。
3. 参照蛋白質のアミノ酸配列が疎水コアの形成に実質的に関与する1又は2以上のコア部分配列と疎水コアの形成に実質的に関与しない1又は2以上のサブ部分配列とに分割された請求の範囲第2項に記載の方法。
4. 参照蛋白質における各アミノ酸残基の側鎖の蛋白質内部への埋没度及び／又は蛋白質表面への露出度の情報と、質問配列の各アミノ酸の疎水性及び／又は親水性の性質とに基づいてマッチングを行う請求の範囲第1項ないし第3項のいずれか1項に記載の方法。
5. 参照蛋白質の1又は2以上のコア部分配列を質問配列上でスライドさせ、該コア部分配列の片端又は両端以外においてはギャップを考慮せずにマッチングを行う請求の範囲第1項ないし第4項のいずれか1項に記載の方法。
6. ギャップが1又は2以上のアミノ酸残基の削除又は付加である請求の範囲第5項に記載の方法。
7. マッチングが以下の工程：
  - (a) 1又は2以上のコア部分配列を質問配列上でスライドさせながら、必要に応じて該コア部分配列の片端又は両端においてはギャップを考慮してマッチングを行う工程（ただし、2以上のコア部分配列を用いる場合には、該コア部分配列は



参照蛋白質のアミノ酸配列での出現順に順番に配置する) ; 及び

(b) 工程(a) に続いて、必要に応じてギャップを考慮しつつサブ部分配列を質問配列上でスライドさせる工程を含む請求の範囲第1項ないし第6項のいずれか1項に記載の方法。

8. 参照蛋白質のアミノ酸残基の側鎖に関する環境情報と質問配列上の対応アミノ酸残基の疎水度パラメータとから計算したスコアに基づいて最適マッチングを選択する、請求の範囲第1項ないし第7項のいずれか1項に記載の方法。

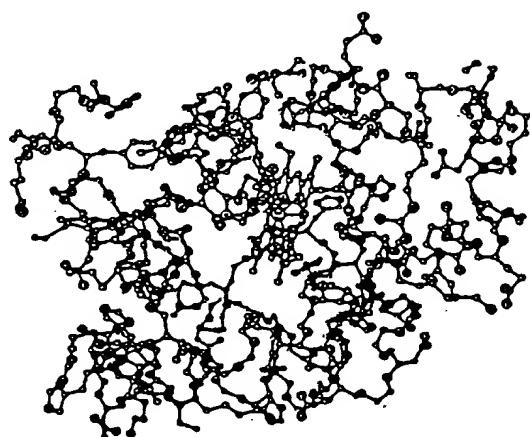
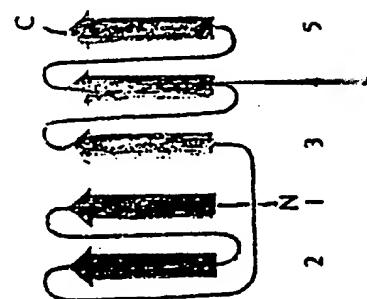
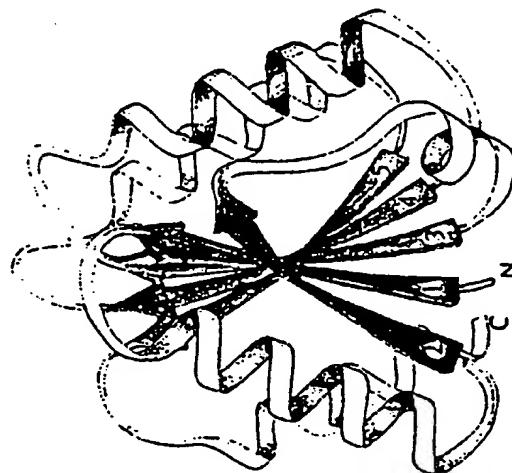
9. 参照蛋白質についての自己マッチングスコアを用いて上記スコアを規格化する工程を含む請求の範囲第8項に記載の方法。

10. さらに質問配列の蛋白質の立体構造を構築する工程を含む請求の範囲第1項ないし第9項のいずれか1項に記載の方法。

11. 立体構造が既知又は推定可能な参照蛋白質のアミノ酸配列に含まれる各アミノ酸残基の側鎖についての環境情報を含むデータベースであって、請求の範囲第1項ないし第10項のいずれか1項の方法に用いるためのデータベース。



第1図



430 Rec'd PCT/PTO 03 APR 2000

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP98/04457

## A. CLASSIFICATION OF SUBJECT MATTER

Int.C1<sup>6</sup> G01N33/68, G06F17/00, G06F15/18, C07K14/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.C1<sup>6</sup> G01N33/68, G06F17/00, G06F15/18, C07K14/00Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-1998  
Kokai Jitsuyo Shinan Koho 1971-1998 Jitsuyo Shinan Toroku Koho 1996-1998

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

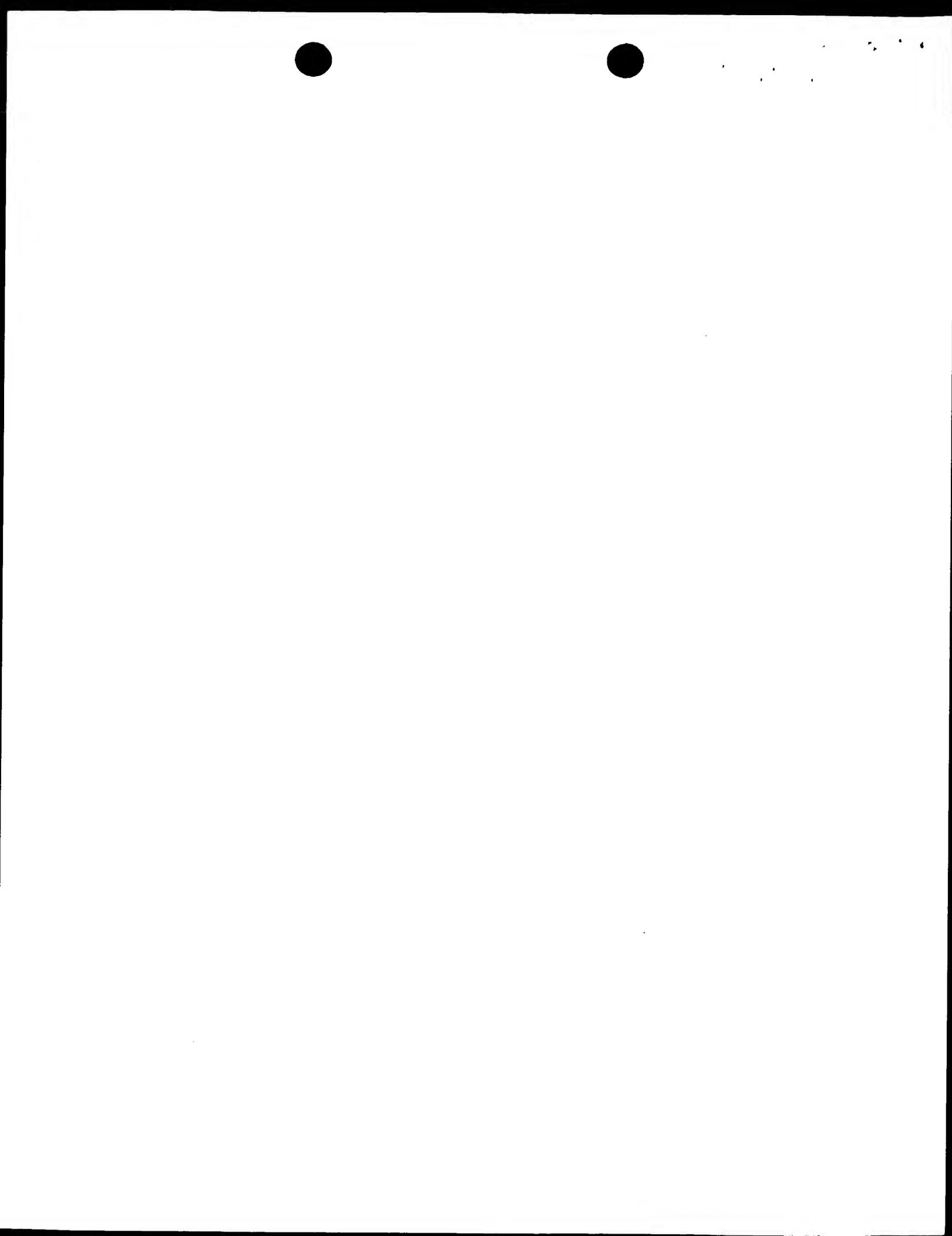
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PA	JP, 10-95796, A (NEC Corp.), 14 April, 1998 (14. 04. 98) (Family: none)	1-10
X	JP, 7-206894, A (Advanced Technology Institute K.K.), 8 August, 1995 (08. 08. 95) (Family: none)	1-10
A	JP, 8-69446, A (NEC Corp.), 12 March, 1996 (12. 03. 96) (Family: none)	1-10
A	JP, 7-105179, A (NEC Corp.), 21 April, 1995 (21. 04. 95) (Family: none)	1-10
A	JP, 7-93286, A (NEC Corp.), 7 April, 1995 (07. 04. 95) (Family: none)	1-10
A	JP, 7-13959, A (NEC Corp.), 17 January, 1995 (17. 01. 95) (Family: none)	1-10

 Further documents are listed in the continuation of Box C.  See patent family annex.

• Special categories of cited documents:	
• "A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
• "E" earlier document but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
• "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
• "O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
• "P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 14 December, 1998 (14. 12. 98)	Date of mailing of the international search report 6 January, 1999 (06. 01. 99)
---	--

Name and mailing address of the ISA/ Japanese Patent Office	Authorized officer
Facsimile No.	Telephone No.

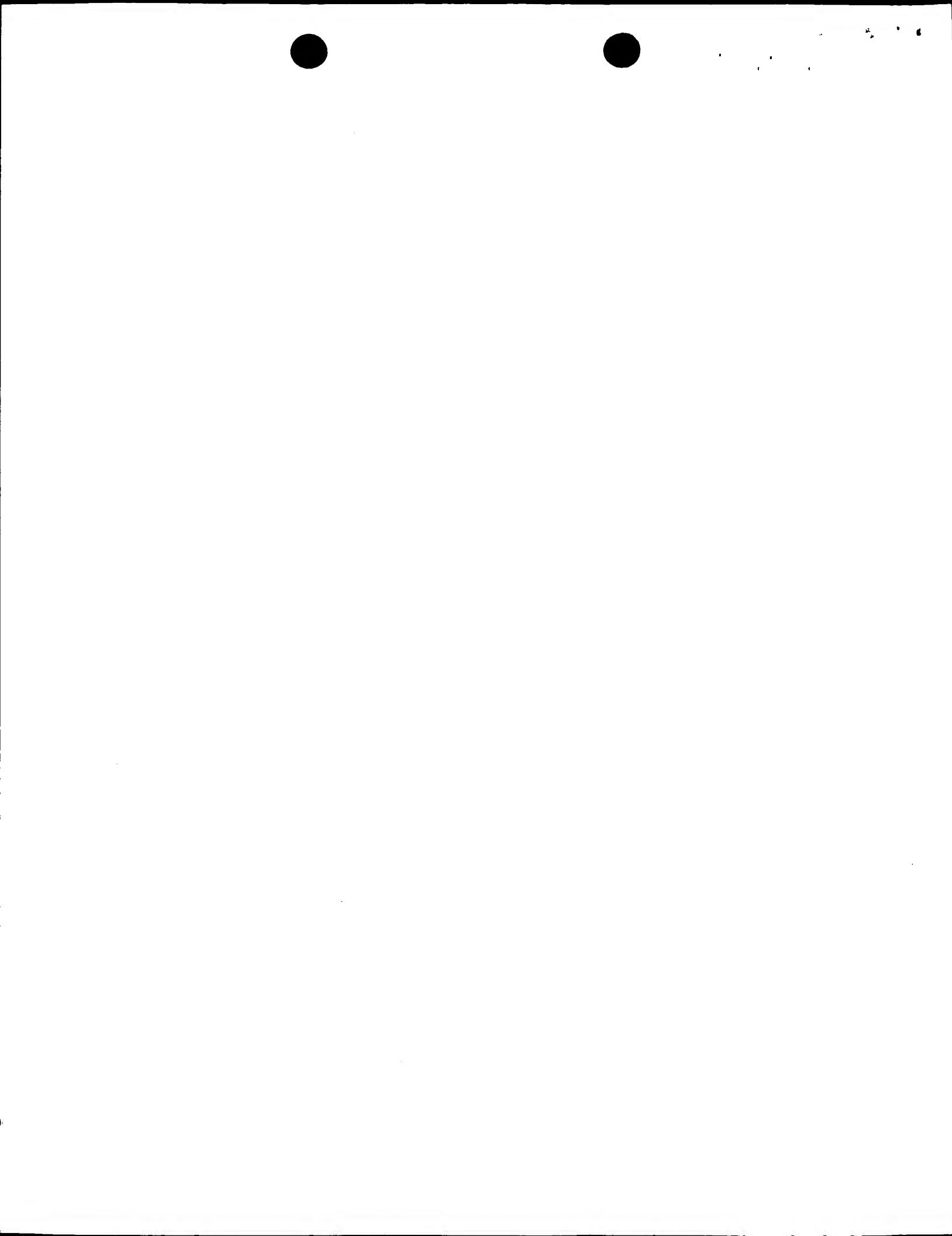


## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/JP98/04457

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP, 5-282383, A (Fujitsu Ltd.), 29 October, 1993 (29. 10. 93) (Family: none)	1-10
A	JP, 4-45781, A (Fujitsu Ltd.), 14 February, 1992 (14. 02. 92) (Family: none)	1-10



A. 発明の属する分野の分類 (国際特許分類 (IPC))  
Int. Cl. G01N33/68, G06F17/00, G06F15/18, C07K14/00

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))  
Int. Cl. G01N33/68, G06F17/00, G06F15/18, C07K14/00

## 最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-1998年
日本国登録実用新案公報	1994-1998年
日本国実用新案登録公報	1996-1998年

## 国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
P A	JP, 10-95796, A (日本電気株式会社) 14. 4月. 1998 (14. 04. 98) (ファミリーなし)	1~10
X	JP, 7-206894, A (アドバンスドテクノロジーインスティチュート株式会社) 8. 8月. 1995 (08. 08. 95) (ファミリーなし)	1~10
A	JP, 8-69446, A (日本電気株式会社) 12. 3月. 1996 (12. 03. 96) (ファミリーなし)	1~10
A	JP, 7-105179, A (日本電気株式会社) 21. 4月. 1995 (21. 04. 95) (ファミリーなし)	1~10

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの  
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
「O」口頭による開示、使用、展示等に言及する文献  
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献  
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
「&」同一パテントファミリー文献

## 国際調査を完了した日

14. 12. 98

## 国際調査報告の発送日

06.01.99

## 国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

## 特許庁審査官 (権限のある職員)

亀田 宏之



2 J 9015

電話番号 03-3581-1101 内線 3252



C (続き) 関連すると認められる文献		関連する 請求の範囲の番号
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	
A	JP, 7-93286, A (日本電気株式会社) 7. 4月. 1995 (07. 04. 95) (ファミリーなし)	1~10
A	JP, 7-13959, A (日本電気株式会社) 17. 1月. 1995 (17. 01. 95) (ファミリーなし)	1~10
A	JP, 5-282383, A (富士通株式会社) 29. 10月. 1993 (29. 10. 93) (ファミリーなし)	1~10
A	JP, 4-45781, A (富士通株式会社) 14. 2月. 1992 (14. 02. 92) (ファミリーなし)	1~10

